# Exact inequalities for sums
# of asymmetric random variables,
# with applications

## Iosif Pinelis

*Department of Mathematical Sciences*
*Michigan Technological University*
*Houghton, Michigan 49931, USA*
*E-mail:* `ipinelis@mtu.edu`

**Abstract:** Let $\mathrm{BS}_1, \ldots, \mathrm{BS}_n$ be independent identically distributed random variables each having the standardized Bernoulli distribution with parameter $p \in (0,1)$. Let $m_*(p) := (1 + p + 2\,p^2)/(2\sqrt{p - p^2 + 4\,p^2})$ if $0 < p \leqslant \frac{1}{2}$ and $m_*(p) := 1$ if $\frac{1}{2} \leqslant p < 1$. Let $m \geqslant m_*(p)$. Let $f$ be such a function that $f$ and $f''$ are nondecreasing and convex. Then it is proved that for all nonnegative numbers $c_1, \ldots, c_n$ one has the inequality

$$\mathsf{E}\, f(c_1 \mathrm{BS}_1 + \cdots + c_n \mathrm{BS}_n) \leqslant \mathsf{E}\, f\left( s^{(m)} \cdot (\mathrm{BS}_1 + \cdots + \mathrm{BS}_n) \right),$$

where $s^{(m)} := \left( \frac{1}{n} \sum_{i=1}^{n} c_i^{2m} \right)^{\frac{1}{2m}}$. The lower bound $m_*(p)$ on $m$ is exact for each $p \in (0,1)$. Moreover, $\mathsf{E}\, f(c_1 \mathrm{BS}_1 + \cdots + c_n \mathrm{BS}_n)$ is Schur-concave in $(c_1^{2m}, \ldots, c_n^{2m})$. A number of related results are presented, including ones for the "symmetric" case.

A number of corollaries are obtained, including upper bounds on generalized moments and tail probabilities of (super)martingales with differences of bounded asymmetry, and also upper bounds on the maximal function of such (super)martingales. It is shown that these results may be important in certain statistical applications.

**AMS 2000 subject classifications:** Primary 60E15, 60G50, 60G42, 60G48, 62F03, 62G25, 62G10, 60G15; secondary 60E05, 62E10, 62G35.
**Keywords and phrases:** supermartingales, martingales, upper bounds, probability inequalities, generalized moments, $t$ statistic, self-normalized sums.

## Contents

## 1. Introduction

Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher random variables (r.v.'s), so that $\mathsf{P}(\varepsilon_i = 1) = \mathsf{P}(\varepsilon_i = -1) = \frac{1}{2}$ for all $i$. Let $Z \sim N(0,1)$. Let $a_1, \ldots, a_n$ be any real numbers such that
$$a_1^2 + \cdots + a_n^2 = 1.$$
The sharp form,
$$\mathsf{E}\, f\,(\varepsilon_1 a_1 + \cdots + \varepsilon_n a_n) \leqslant \mathsf{E}\, f(Z), \qquad (1.1)$$
of Khinchin's inequality [21] for $f(x) \equiv |x|^p$ was proved by Whittle (1960) [37] for $p \geqslant 3$ and Haagerup (1982) [17] for $p \geqslant 2$.

For $f(x) \equiv e^{\lambda x}$ ($\lambda \geqslant 0$), inequality (1.1) follows from Hoeffding (1963) [18], whence

$$\mathsf{P}\,(\varepsilon_1 a_1 + \cdots + \varepsilon_n a_n \geqslant x) \leqslant \inf_{\lambda \geqslant 0} \frac{\mathsf{E}\, e^{\lambda Z}}{e^{\lambda x}} = e^{-x^2/2} \quad \forall x \geqslant 0. \qquad (1.2)$$

Since $\mathsf{P}(Z \geqslant x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$ as $x \to \infty$, a factor $\asymp \frac{1}{x}$ is "missing" here. The apparent cause of this deficiency is that the class of exponential moment functions $f(x) \equiv e^{\lambda x}$ ($\lambda \geqslant 0$) is too small (and so is the class of the power functions $f(x) \equiv |x|^p$).

For all $\alpha \geqslant 0$, consider the following much richer classes of functions $f \colon \mathbb{R} \to \mathbb{R}$:
$$\mathcal{H}_+^\alpha := \{f \colon f(x) = \textstyle\int_{-\infty}^\infty (x-t)_+^\alpha\, \mu(dt) \quad \forall u \in \mathbb{R}\}, \qquad (1.3)$$
where $\mu \geqslant 0$ is a Borel measure, $x_+ := \max(0, x)$ and $x_+^\alpha := (x_+)^\alpha$ for $x \in \mathbb{R}$, $0^0 := 0$; however, the subscript $_+$ will have a different meaning when used with functions or classes of functions (as, for example, in the symbol $\mathcal{H}_+^\alpha$).

It is easy to see [29, Proposition 1(ii)] that

$$0 \leqslant \beta < \alpha \quad \text{implies} \quad \mathcal{H}_+^\alpha \subseteq \mathcal{H}_+^\beta. \qquad (1.4)$$

**Proposition 1.1.** [31] *For natural $\alpha$, one has $f \in \mathcal{H}_+^\alpha$ if and only if $f$ has finite derivatives $f^{(0)} := f, f^{(1)} := f', \ldots, f^{(\alpha-1)}$ on $\mathbb{R}$ such that $f^{(\alpha-1)}$ is convex on $\mathbb{R}$ and $f^{(j)}(-\infty+) = 0$ for $j = 0, 1, \ldots, \alpha - 1$.*

It follows from Proposition 1.1 that, for every $t \in \mathbb{R}$, every $\beta \geqslant \alpha$, and every $\lambda > 0$, the functions $u \mapsto (u-t)_+^\beta$ and $u \mapsto e^{\lambda(u-t)}$ belong to $\mathcal{H}_+^\alpha$.

Eaton (1970) [12] proved the Khinchin-Whittle-Haagerup inequality (1.1) for a class of moment functions, which essentially coincides with the class $\mathcal{H}_+^3$; see

[27, Proposition A.1]. Based on asymptotics, numerics, and a certain related inequality, Eaton (1974) [13] conjectured that the mentioned moment comparison inequality of his implies that

$$\mathsf{P}\left(\varepsilon_1 a_1 + \cdots + \varepsilon_n a_n \geqslant x\right) \leqslant \frac{2e^3}{9} \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \quad \forall x > \sqrt{2}.$$

Pinelis (1994) [27] proved the following improvement of this conjecture:

$$\mathsf{P}\left(\varepsilon_1 a_1 + \cdots + \varepsilon_n a_n \geqslant x\right) \leqslant \frac{2e^3}{9} \mathsf{P}(Z \geqslant x) \quad \forall x \in \mathbb{R}. \tag{1.5}$$

It was realized in Pinelis (1998) [28] that the reason why it is possible to extract tail comparison inequality (1.5) from the Khinchin-Eaton moment comparison inequality (1.1) for $f \in \mathcal{H}_+^3$ is that the tail function $x \mapsto \mathsf{P}(Z \geqslant x)$ is log-concave. This realization resulted in a general device, which allows one to extract an optimal tail comparison inequality from an appropriate moment comparison inequality. The following is a special case of Theorem 4 of Pinelis (1999) [29]; see also Theorem 3.11 of Pinelis (1998) [28].

**Theorem 1.2.** *Suppose that $0 \leqslant \beta \leqslant \alpha$, $\xi$ and $\eta$ are real-valued r.v.'s, and the tail function $u \mapsto \mathsf{P}(\eta \geqslant u)$ is log-concave on $\mathbb{R}$. Then the comparison inequality*

$$\mathsf{E}\, f(\xi) \leqslant \mathsf{E}\, f(\eta) \quad \text{for all } f \in \mathcal{H}_+^\alpha \tag{1.6}$$

*implies*

$$\mathsf{E}\, f(\xi) \leqslant c_{\alpha,\beta}\, \mathsf{E}\, f(\eta) \quad \text{for all } f \in \mathcal{H}_+^\beta \tag{1.7}$$

*and, in particular, for all real $x$,*

$$\mathsf{P}(\xi \geqslant x) \leqslant \inf_{f \in \mathcal{H}_+^\alpha} \frac{\mathsf{E}\, f(\eta)}{f(x)} \tag{1.8}$$

$$= B_{\boldsymbol{opt}}(x) := \inf_{t \in (-\infty, x)} \frac{\mathsf{E}(\eta - t)_+^\alpha}{(x - t)^\alpha} \tag{1.9}$$

$$\leqslant \min\left(c_{\alpha,0}\, \mathsf{P}(\eta \geqslant x), \inf_{h>0} e^{-hx}\, \mathsf{E}\, e^{h\eta}\right), \tag{1.10}$$

*where*

$$c_{\alpha,\beta} := \frac{\Gamma(\alpha+1)(e/\alpha)^\alpha}{\Gamma(\beta+1)(e/\beta)^\beta} \tag{1.11}$$

*for $\beta > 0$; $c_{\alpha,0} := \Gamma(\alpha+1)(e/\alpha)^\alpha$. Moreover, the constant $c_{\alpha,\beta}$ is the best possible in (1.7) and (1.10).*

A similar result for the case when $\alpha = 1$ and $\beta = 0$ is contained in the book by Shorack and Wellner (1986) [36], pages 797–799.

Note that $c_{\alpha,0} \sim \sqrt{2\pi\alpha}$ and $c_{\alpha,\beta} \sim \sqrt{\alpha/\beta}$ as $\alpha, \beta \to \infty$.

*Remark* 1.3. As follows from [28, Remark 3.13], a useful point is that the requirement of the log-concavity of the tail function $q(u) := \mathsf{P}(\eta \geqslant u)$ in Theorem 1.2 can be relaxed by replacing $q$ with any (e.g., the least) log-concave majorant of $q$. However, then the optimality of $c_{\alpha,\beta}$ is not guaranteed.

Detailed studies of various cases and aspects of the optimal bound $B_{\mathsf{opt}}(x)$ in (1.9) were presented in [11, 28, 7].

Note that $c_{3,0} = 2e^3/9$, which is the constant factor in (1.5). Bobkov, Götze, and Houdré (BGH) (2001) [9] obtained a simpler proof of inequality (1.5), but with a constant factor $\approx 12.0$ in place of $2e^3/9 \approx 4.46$. In [34] the BGH method was modified to obtain a version of (1.5) with a constant factor $\approx 3.22$, which is $\approx 1.01$ times the least possible constant factor in (1.5). Edelman [14] proposed inequality $\mathsf{P}(S_n \geqslant x) \leqslant \mathsf{P}(Z \geqslant x - 1.5/x)$ for all $x > 0$, but his proof appears to have a gap. A more precise upper bound, with $\ln c_{3,0} = 1.495\ldots$ in place of 1.5, was recently shown [33] to be a rather easy corollary of (1.5).

*Remark* 1.4. One also has two kinds of multi-dimensional analogues of (1.1) and (1.5). One kind is represented by [27, Theorems 2.3 and 2.4]. The other kind is based on the dimensionality reduction device given in [30]. Indeed, Remarks in [13] imply (cf. the proof of Lemma 3.2 in [27]) that, for any even function $f$ in class $\mathcal{F}^3$ (which contains $\mathcal{H}_+^3$ and is defined by (2.14) below), the function $[0, \infty) \ni u \mapsto f(\sqrt{u})$ is convex. Therefore, by [30, Theorem 2.1],

$$\mathsf{E} f\left(\|\varepsilon_1 \mathbf{x}_1 + \cdots + \varepsilon_n \mathbf{x}_n\|\right) \leqslant \mathsf{E} f(|Z|) \quad \text{for all even } f \text{ in } \mathcal{F}^3, \qquad (1.12)$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are any non-random vectors in a Hilbert space $(H, \|\cdot\|)$ such that $\|\mathbf{x}_1\|^2 + \cdots + \|\mathbf{x}_n\|^2 = 1$. It follows that (1.12) holds for all functions $f$ given by $f(x) := \mathsf{E} g(\varepsilon x)$ $(x \in \mathbb{R})$, where $g \in \mathcal{H}_+^3$ and $\varepsilon$ is a Rademacher r.v. Hence, by Theorem 1.2,

$$\mathsf{P}\left(\|\varepsilon_1 \mathbf{x}_1 + \cdots + \varepsilon_n \mathbf{x}_n\| \geqslant x\right) \leqslant \frac{2e^3}{9} \, \mathsf{P}(|Z| \geqslant x) \quad \forall x \in \mathbb{R}, \qquad (1.13)$$

More generally, in view of a result by Hunt [20], inequalities (1.1), (1.5), (1.12), and (1.13) hold if $\varepsilon_1, \ldots, \varepsilon_n$ are replaced there by any independent zero-mean r.v.'s $\eta_1, \ldots, \eta_n$ such that $|\eta_i| \leqslant 1$ almost surely (a.s.) for all $i$.

Pinelis (1999) [29] also obtained the "discrete" improvement of (1.5):

$$\mathsf{P}\left(\varepsilon_1 a_1 + \cdots + \varepsilon_n a_n \geqslant x\right) \leqslant \frac{2e^3}{9} \, \mathsf{P}\left(\frac{1}{\sqrt{n}}(\varepsilon_1 + \cdots + \varepsilon_n) \geqslant x\right) \qquad (1.14)$$

for all values $x$ of r.v. $\frac{1}{\sqrt{n}}(\varepsilon_1 + \cdots + \varepsilon_n)$.

Such results can be, and have been, extended in several different directions. In what follows, let $(S_0, S_1, \ldots)$ be a supermartingale relative to a nondecreasing sequence of $\sigma$-algebras $(H_{\leqslant 0}, H_{\leqslant 1}, \ldots)$, with $S_0 \leqslant 0$ a.s. and differences

$$X_i := S_i - S_{i-1}, \quad i = 1, 2, \ldots.$$

The following normal domination statement is one of the main results of [32].

**Theorem 1.5.** [32] *Suppose that for every $i = 1, 2, \ldots$ there exist $H_{\leqslant(i-1)}$-measurable r.v.'s $A_{i-1}$ and $B_{i-1}$ and a positive real number $c_i$ such that*

$$-A_{i-1} \leqslant X_i \leqslant B_{i-1} \quad and \qquad (1.15)$$

$$\tfrac{1}{2}(A_{i-1} + B_{i-1}) \leqslant c_i \qquad (1.16)$$

*a.s. Then for all $f \in \mathcal{H}_+^5$ and all $n = 1, 2, \ldots$*

$$\mathsf{E}\, f(S_n) \leqslant \mathsf{E}\, f(s\sqrt{n}Z), \tag{1.17}$$

*where*

$$s := s^{(1)} := \sqrt{\frac{c_1^2 + \cdots + c_n^2}{n}}. \tag{1.18}$$

Note that inequality (1.17) for the smaller class of exponential functions in place of the class $\mathcal{H}_+^5$ is due to Hoeffding [18].

By virtue of Theorem 1.2, one has the following corollary.

**Corollary 1.6.** [32] *Under the conditions of Theorem 1.5, for all $\beta \in [0, 5]$, all $f \in \mathcal{H}_+^\beta$, and all $n = 1, 2, \ldots$*

$$\mathsf{E}\, f(S_n) \leqslant c_{5,\beta}\, \mathsf{E}\, f(s\sqrt{n}Z). \tag{1.19}$$

*In particular, for all real $x$,*

$$\mathsf{P}(S_n \geqslant x) \leqslant \inf_{f \in \mathcal{H}_+^5} \frac{\mathsf{E}\, f(s\sqrt{n}Z)}{f(x)} \tag{1.20}$$

$$= \inf_{t \in (-\infty, x)} \frac{\mathsf{E}(s\sqrt{n}Z - t)_+^5}{(x - t)^5} \tag{1.21}$$

$$\leqslant \min\left( c_{5,0}\, \mathsf{P}(s\sqrt{n}Z \geqslant x), \inf_{h>0} e^{-hx}\, \mathsf{E}\, e^{hs\sqrt{n}Z} \right) \tag{1.22}$$

$$= \min\left( c_{5,0}\, \mathsf{P}\left( Z \geqslant \tfrac{x}{s\sqrt{n}} \right), \exp\left( -\tfrac{x^2}{2ns^2} \right) \right). \tag{1.23}$$

The upper bound (1.23) – but with a constant factor greater than 427 in place of $c_{5,0} = 5!(e/5)^5 = 5.699\ldots$ was obtained in Bentkus (2001) [1] for the case when $(S_i)$ is a martingale. (In this case, Bentkus was using direct methods, rather than a generalized moment comparison inequality such as (1.17).) The large value, 427, of the constant factor renders the bound in [1] hardly usable in statistics. Indeed, the upper bound $427\,\mathsf{P}\left( Z \geqslant \tfrac{x}{s\sqrt{n}} \right)$ improves the Hoeffding bound $\exp\left( -\tfrac{x^2}{2ns^2} \right)$ only when $\tfrac{x}{s\sqrt{n}} > 170$, in which case (in view of (1.23)) one has $\mathsf{P}(S_n \geqslant x) < c_{5,0}\, \mathsf{P}(Z \geqslant 170) < 10^{-6200}$.

As shown in [32], Theorem 1.5 and Corollary 1.6 are well suited in order to obtain the most precise presently known bounds for the measure concentration phenomenon in terms of separately-Lipschitz (or, equivalently, $\ell^1$-Lipschitz) functions on product spaces.

Theorem 1.5 can be further improved, as follows.

**Theorem 1.7.** [32] *Suppose that for every $i = 1, 2, \ldots$ there exist a positive $H_{\leqslant(i-1)}$-measurable r.v. $B_{i-1}$ and a positive real number $\hat{s}_i$ such that*

$$X_i \leqslant B_{i-1} \quad and \tag{1.24}$$

$$\frac{1}{2}\left( B_{i-1} + \frac{\mathsf{Var}_{i-1} X_i}{B_{i-1}} \right) \leqslant \hat{s}_i \tag{1.25}$$

*a.s. Here and elsewhere, we let* $\mathsf{Var}_j$ *stand for conditional variance given* $H_{\leqslant j}$. *Then one has all the inequalities* (1.17) *and* (1.19)–(1.23), *only with s replaced by*

$$\hat{s} := \sqrt{\frac{\hat{s}_1^2 + \cdots + \hat{s}_n^2}{n}}. \tag{1.26}$$

The set of conditions (1.24)–(1.25) is equivalent to

$$X_i \leqslant B_{i-1} \quad \text{and} \quad \sigma_*(B_{i-1}, \mathsf{Var}_{i-1} X_i) \leqslant \hat{s}_i$$

a.s., where

$$\sigma_*(b_0, c^2) := \frac{1}{2} \inf_{b \geqslant b_0} \left( b + \frac{c^2}{b} \right) = \min \left( c \vee b_0, \frac{1}{2} \left( b_0 + \frac{c^2}{b_0} \right) \right)$$

$$= \begin{cases} c & \text{if } c \geqslant b_0, \\ \frac{1}{2} \left( b_0 + \frac{c^2}{b_0} \right) & \text{if } c < b_0, \end{cases}$$

for positive $c$ and $b_0$. This follows simply because the inequalities $X_i \leqslant B_{i-1}$ and $b \geqslant B_{i-1}$ imply $X_i \leqslant b$.

Thus, in the case when $\mathsf{Var}_{i-1} X_i < B_{i-1}^2$ a.s., conditions (1.24)–(1.25) represent an improvement of condition $B_{i-1}^2 \vee \mathsf{Var}_{i-1} X_i \leqslant \hat{s}_i^2$ a.s., imposed in [3, 4]. In a certain variety of cases, this improvement may be even more significant than the improvement in the constant factor from 427 to 5.699 ... *before* the probability sign.

Moreover, it can be shown that the function $\sigma_*(\cdot, \cdot)$ of the pair $(B_{i-1}, \mathsf{Var}_{i-1} X_i)$, which is in effect used in Theorem 1.7 is nearly optimal as far as the normal domination is concerned.

However, it can also be seen that even the best possible normal domination may be inadequate if the asymmetry of the random summands $X_i$ is significant or if $n$ is not large. In such a case, one may try to use binomial domination instead of normal, as in the following theorem, which is a straighforward corollary of results of [2] (or [4]).

**Theorem 1.8.** [2, 4] *Suppose that for every* $i = 1, 2, \ldots$ *there exist non-random constants* $b_i > 0$ *and* $c_i > 0$ *such that*

$$X_i \leqslant b_i \quad \text{and} \tag{1.27}$$
$$\mathsf{Var}_{i-1} X_i \leqslant c_i^2 \tag{1.28}$$

*a.s. Then, for all* $n = 1, 2, \ldots$,

$$\mathsf{E} f(S_n) \leqslant \mathsf{E} f(T_n) \quad \forall f \in \mathcal{H}_+^2, \quad \text{where} \tag{1.29}$$
$$T_n := Z_1 + \cdots + Z_n \tag{1.30}$$

*and* $Z_1, \ldots, Z_n$ *are i.i.d. r.v.'s such that each* $Z_i$ *takes on only two values, one of which is*

$$b := \max_i b_i, \tag{1.31}$$

*and satisfies the conditions*

$$\mathsf{E}\, Z_i = 0 \quad and \quad \mathsf{Var}\, Z_i = c^2, \quad where$$

$$c := \left( \frac{1}{n} \sum_{i=1}^{n} c_i^2 \right)^{1/2}; \tag{1.32}$$

*that is,* $\quad \mathsf{P}(Z_i = b) = \dfrac{c^2}{b^2 + c^2} \quad and \quad \mathsf{P}\left( Z_i = -\dfrac{c^2}{b} \right) = \dfrac{b^2}{b^2 + c^2}.$

Based on this result, the tail comparison inequality

$$\mathsf{P}(S_n \geqslant y) \leqslant c_{2,0}\, \mathsf{P}^{\mathsf{Lin,LC}}(T_n \geqslant y + \tfrac{h}{2}) \quad \forall y \in \mathbb{R} \tag{1.33}$$

was obtained in [31], where $c_{2,0} = e^2/2 = 3.694\ldots$ (in accordance with (1.11)), $h := b + c^2/b$, and the function $y \mapsto \mathsf{P}^{\mathsf{Lin,LC}}(T_n \geqslant y)$ is the least log-concave majorant of the linear interpolation of the tail function $y \mapsto \mathsf{P}(T_n \geqslant y)$ over the lattice of all points of the form $nb+kh$ ($k \in \mathbb{Z}$). Tail comparison inequality (1.33) is a substantial improvement of the corresponding inequality in [2, Theorem 1] and [4, Theorem 1.1].

Yet, while the "variance"-averaging given by (1.32) is nice, the extreme kind of upper-bound-averaging (1.31) seems very crude.

In this paper, another approach to the problem of asymmetry is presented. Here we provide binomial upper bounds on generalized moments and tail probabilities for $S_n$ assuming that certain *indices of asymmetry* of the $X_i$'s (rather than the $X_i$'s themselves) are *uniformly* bounded from above. This assumption of bounded asymmetry (in contrast with the uniform boundedness) of the $X_i$'s is natural in certain statistical applications; see Subsection 3.2.

The results of [6] for the symmetric case can be similarly complemented, with condition (1.31) of the uniform boundedness of the ranges now replaced by a condition of uniform boundedness of the kurtoses of the $X_i$'s; see Remark 3.6.

## 2. Statements of basic results and discussion

Let $\mathsf{BS}(p)$ denote the standardized Bernoulli distribution with parameter $p$: for a r.v. $\mathsf{BS}$ let, by definition,

$$\mathsf{BS} \sim \mathsf{BS}(p) \iff \mathsf{P}\left( \mathsf{BS} = \sqrt{\tfrac{q}{p}} \right) = p = 1 - \mathsf{P}\left( \mathsf{BS} = -\sqrt{\tfrac{p}{q}} \right),$$

where

$$q := 1 - p \quad and \quad 0 < p < 1;$$

thus, $\mathsf{BS}(p)$ is a two-point zero-mean unit-variance distribution. In particular, $\mathsf{BS}(\frac{1}{2})$ is the distribution of a Rademacher r.v. $\varepsilon$, with $\mathsf{P}(\varepsilon = \pm 1) = \frac{1}{2}$.

Let $\mathcal{C}^2$ denote the class of all twice continuously differentiable functions $f\colon \mathbb{R} \to \mathbb{R}$. Consider the following class of functions:

$$\mathcal{F}_+^3 := \{f \in \mathcal{C}^2 \colon f \text{ and } f'' \text{ are nondecreasing and convex}\}. \tag{2.1}$$

An equivalent definition would be given by the formula

$$\mathcal{F}_+^3 = \{f \in \mathcal{C}^2 \colon f,\, f',\, f'',\, \text{and } f''' \text{ are nondecreasing}\},$$

where $f'''$ denotes the right derivative of the convex function $f''$.

For example, functions $x \mapsto a + b\, x + c\, (x - t)_+^\alpha$ and $x \mapsto a + b\, x + c\, e^{\lambda x}$ belong to $\mathcal{F}_+^3$ for all $a \in \mathbb{R}$, $b \geqslant 0$, $c \geqslant 0$, $t \in \mathbb{R}$, $\alpha \geqslant 3$, and $\lambda \geqslant 0$.

*Remark* 2.1. If a function $f \colon \mathbb{R} \to \mathbb{R}$ is convex and a r.v. $X$ has a finite expectation, then, by Jensen's inequality, $\mathsf{E}\, f(X)$ always exists in $(-\infty, \infty]$. This remark will be used in this paper (sometimes tacitly) for functions $f$ in the class $\mathcal{F}_+^3$, as well as for other convex functions.

Throughout the paper, unless indicated otherwise, the following notation/assumptions will be used:

$$m \in [1, \infty), \quad p \in (0, 1), \quad q = 1 - p, \quad \text{and} \quad \mathsf{BS}_1, \dots, \mathsf{BS}_n \overset{\text{i.i.d.}}{\sim} \mathsf{BS}(p). \quad (2.2)$$

Introduce also

$$m_*(p) := \begin{cases} \dfrac{1 + p + 2\, p^2}{2\left(\sqrt{p - p^2} + 2\, p^2\right)} & \text{if} \quad 0 < p \leqslant \tfrac{1}{2}, \\ 1 & \text{if} \quad \tfrac{1}{2} \leqslant p < 1. \end{cases} \quad (2.3)$$

Later it will be clear that $m_*(p)$ increases from 1 to $\infty$ as $p$ decreases from $\tfrac{1}{2}$ to 0 (see the proof of Lemma 4.1.17).

Of the main results of this paper, the following one is perhaps the easiest to state (but not to prove).

**Theorem 2.2.** *For any real number*

$$m \geqslant m_*(p),$$

*all $f \in \mathcal{F}_+^3$, all natural $n$, and all nonnegative numbers $c_1, \dots, c_n$, one has*

$$\mathsf{E}\, f(c_1 \mathsf{BS}_1 + \dots + c_n \mathsf{BS}_n) \leqslant \mathsf{E}\, f\left(s^{(m)} \cdot (\mathsf{BS}_1 + \dots + \mathsf{BS}_n)\right), \quad (2.4)$$

$$\text{where} \quad s^{(m)} := \left(\frac{1}{n} \sum_{i=1}^n c_i^{2m}\right)^{\frac{1}{2m}}. \quad (2.5)$$

*Moreover, the lower bound $m_*(p)$ on $m$ is exact for each $p \in (0, 1)$.*

The proofs are deferred to Section 4.

*Remark* 2.3. The general restriction $m \geqslant 1$ in (2.2) is quite natural. Indeed, if inequality (2.4) held for some $m \in (0, 1)$ then, taking $c_1 = 1$, $c_2 = \dots = c_n = 0$, and letting $n \to \infty$, one would have, by the central limit theorem, the inequality $\mathsf{E}\, f(\mathsf{BS}_1) \leqslant f(0)$ for all $f \in \mathcal{F}_+^3$, which is false even for $f(x) \equiv e^x$ or $f(x) \equiv x_+^3$.

Here is a generalization of Theorem 2.2:

**Theorem 2.4.** *Let $X_1, \ldots, X_n$ be independent r.v.'s such that for every $i \in \{1, \ldots, n\}$*

$$\mathsf{E}\, X_i \leqslant 0 \quad and \quad -a_i \leqslant X_i \leqslant b_i \ a.s.,$$

*where $a_i$ and $b_i$ are positive numbers such that*

$$\frac{b_i}{a_i} \leqslant \frac{q}{p}. \tag{2.6}$$

*Then, for any real number $m \geqslant m_*(p)$ and all $f \in \mathcal{F}_+^3$, one has the inequality*

$$\mathsf{E}\, f(X_1 + \cdots + X_n) \leqslant \mathsf{E}\, f\left(s^{(m)} \cdot (\mathrm{BS}_1 + \cdots + \mathrm{BS}_n)\right), \quad where \tag{2.7}$$

$$s^{(m)} := \left(\frac{1}{n}\sum_{i=1}^n (a_i b_i)^m\right)^{\frac{1}{2m}}.$$

*Moreover, the lower bound $m_*(p)$ on $m$ is exact for each $p \in (0,1)$.*

Condition (2.6) may be referred to as a bounded-asymmetry coundition. Theorem 2.4 can be easily extended to (super)martingales.

**Theorem 2.5.** *Suppose that for every $i \in \{1, \ldots, n\}$ one has (1.15) with positive $A_{i-1}$ and $B_{i-1}$ such that*

$$\sqrt{A_{i-1}\, B_{i-1}} \leqslant c_i \quad and \tag{2.8}$$

$$\frac{B_{i-1}}{A_{i-1}} \leqslant \frac{q}{p} \tag{2.9}$$

*a.s., where $c_i$ is a non-random number. Then, for any real number*

$$m \geqslant m_*(p) \tag{2.10}$$

*and $f \in \mathcal{F}_+^3$, one has the inequality*

$$\mathsf{E}\, f(S_n) \leqslant \mathsf{E}\, f\left(s^{(m)} \cdot (\mathrm{BS}_1 + \cdots + \mathrm{BS}_n)\right), \tag{2.11}$$

*where $s^{(m)}$ is defined by (2.5). Moreover, the lower bound $m_*(p)$ on $m$ is exact for each $p \in (0,1)$.*

One should compare (2.8) and (2.5) with (1.16) and (1.18). If $X_{a,b}$ stands for a zero-mean r.v. taking on values in the set $\{-a, b\}$ for some positive $a$ and $b$, then obviously the half-range $\frac{1}{2}(a+b)$ of $X_{a,b}$ is no less than its standard deviation $\sqrt{ab}$. That is, (1.16) is more restrictive than (2.8). On the other hand, one has the inequality $s^{(m)} \geqslant s^{(1)}$ for $m \geqslant 1$. Moreover, the greater the uniform bound $\frac{q}{p}$ on asymmetry in (2.9) is, the greater $m$ must be according to (2.10) and hence the more pronounced the inequality $s^{(m)} \geqslant s^{(1)}$ will be. Yet, it will be demonstrated elsewhere that, overall, (2.8) and (2.5) work better in certain important statistical applications than (1.16) and (1.18). Note also that one

can choose the "ideal" value $m = 1$ whenever the asymmetry index $\frac{q}{p}$ does not exceed 1, that is, whenever the $X_i$'s are not skewed to the right.

Recall the definition of the *Schur majorizarion*: for $\mathbf{a} := (a_1, \ldots, a_n)$ and $\mathbf{b} := (b_1, \ldots, b_n)$ in $\mathbb{R}^n$, $\mathbf{a} \succcurlyeq \mathbf{b}$ means that $a_1 + \cdots + a_n = b_1 + \cdots + b_n$ and $a_{[1]} + \cdots + a_{[j]} \geqslant b_{[1]} + \cdots + b_{[j]}$ for all $j \in \{1, \ldots, n\}$, where $a_{[1]} \geqslant \cdots \geqslant a_{[n]}$ are the ordered numbers $a_1, \ldots, a_n$, from the largest to the smallest. Recall also that a function $\mathcal{Q} \colon [0, \infty)^n \to \mathbb{R}$ is referred to as *Schur-concave* if it reverses the Schur majorization: for any $\mathbf{a}$ and $\mathbf{b}$ in $[0, \infty)^n$ such that $\mathbf{a} \succcurlyeq \mathbf{b}$, one has $\mathcal{Q}(\mathbf{a}) \leqslant \mathcal{Q}(\mathbf{b})$.

Theorems 2.2, 2.4, and 2.5 are contained in

**Theorem 2.6.** *The following statements are equivalent to one another.*

(I) $m \geqslant m_*(p)$.

(II) *For all $f \in \mathcal{F}_+^3$, all natural $n \geqslant 2$, and all nonnegative numbers $c_1, \ldots, c_n$, one has* (2.4).

(III) *The same as item* (II), *but only for $n = 2$.*

(IV) *For every natural $n \geqslant 2$ and every function $f \in \mathcal{F}_+^3$, the function*

$$[0, \infty)^n \ni (a_1, \ldots, a_n) \longmapsto \mathsf{E}\, f(a_1^{1/(2m)} \mathrm{BS}_1 + \cdots + a_n^{1/(2m)} \mathrm{BS}_n) \quad (2.12)$$

*is Schur-concave.*

(V) *The same as item* (IV), *but only for $n = 2$.*

(VI) *Let the $X_i$'s, $a_i$'s, $b_i$'s, and $s^{(m)}$ be as in the statement of Theorem 2.4. Then one has* (2.7) *for all natural $n$.*

(VII) *Let the $X_i$'s, $A_{i-1}$'s, $B_{i-1}$'s, $c_i$'s, and $s^{(m)}$ be as in the statement of Theorem 2.5. Then one has* (2.11) *for all natural $n$.*

The special case of statement (IV) of Theorem 2.6 with $p = \frac{1}{2}$ and $m = 1$ is essentially the mentioned result due to Whittle [37] and Eaton [12].

From the "right-tail" Theorem 2.6, one can deduce its left-tail and two-tail analogues. Appropriate left-tail and two-tail counterparts of $\mathcal{F}_+^3$ are the following classes of functions:

$$\mathcal{F}_-^3 := \{f \in \mathcal{C}^2 \colon f \text{ and } f'' \text{ are nonincreasing and convex}\} \quad (2.13)$$
$$= \{f \colon \exists g \in \mathcal{F}_+^3 \; \forall x \in \mathbb{R} \; f(x) = g(-x)\} \quad \text{and}$$
$$\mathcal{F}^3 := \{f \in \mathcal{C}^2 \colon f \text{ and } f'' \text{ are convex}\}. \quad (2.14)$$

**Theorem 2.7.** *Theorem 2.6 holds with $\mathcal{F}_-^3$ in place of $\mathcal{F}_+^3$ if the restrictions* (i) $m \geqslant m_*(p)$, (ii) $\mathsf{E}\, X_i \leqslant 0$, (iii) $\frac{b_i}{a_i} \leqslant \frac{q}{p}$, (iv) $(S_0, \ldots, S_n)$ *is a supermartingale with $S_0 \leqslant 0$, and* (v) $\frac{B_{i-1}}{A_{i-1}} \leqslant \frac{q}{p}$ *in Theorem 2.6 are replaced, respectively, with the following:*

(i) $m \geqslant m_*(q)$,

(ii) $\mathsf{E}\, X_i \geqslant 0$,

(iii) $\frac{a_i}{b_i} \leqslant \frac{p}{q}$,

(iv) $(S_0, \dots, S_n)$ *is a submartingale with $S_0 \geqslant 0$, and*

(v) $\dfrac{A_{i-1}}{B_{i-1}} \leqslant \dfrac{p}{q}$.

This "left-tail" analogue is a trivial corollary of Theorem 2.6.

The "two-tail" analogue of Theorem 2.6 is more difficult to prove. It relies in part on Proposition 2.9 below, preceded by the following definition.

**Definition 2.8.** Let us say that a sequence of functions $(f_n)$ in $\mathcal{C}^2$ converges to a function $f$ in $\mathcal{C}^2$ and write $f_n \to f$ (as $n \to \infty$) if $f_n(x) \uparrow f(x)$ and $f_n''(x) \to f''(x)$ for all real $x$. (This stronger notion of convergence will make it easier to verify the convergence of relevant expected values; also, it naturally provides for the relevant classes of functions to be closed.)

For any subset $\mathcal{A}$ of $\mathcal{C}^2$, its *closure* – denoted here by $\mathrm{cl}\,\mathcal{A}$ – will be understood here simply as the set of the limits of all sequences in $\mathcal{A}$ that are convergent in $\mathcal{C}^2$. Obviously, $\mathrm{cl}\,\mathcal{A} \supseteq \mathcal{A}$, for all $\mathcal{A} \subseteq \mathcal{C}^2$.

Obviously, the "two-tail" class $\mathcal{F}^3$ contains both "one-tail" classes $\mathcal{F}_+^3$ and $\mathcal{F}_-^3$. The more informative relation of $\mathcal{F}^3$ to $\mathcal{F}_+^3$ and $\mathcal{F}_-^3$ (on which the proof of Theorem 2.10 below is partly based) is given by

**Proposition 2.9.** *One has $\mathcal{F}^3 = \mathrm{cl}\,\mathcal{G}^3$, where*

$$\mathcal{G}^3 := \{f \in \mathcal{C}^2 \colon \exists c \geqslant 0 \; \exists f_+ \in \mathcal{F}_+^3 \; \exists f_- \in \mathcal{F}_-^3 \; \forall x \in \mathbb{R}$$
$$f(x) = c\,x^2/2 + f_+(x) + f_-(x)\}. \quad (2.15)$$

*However, $\mathcal{F}^3 \neq \mathcal{G}^3$.*

For example, functions $x \mapsto a + b\,x + c\,x^2 + d\,|x - t|^\alpha$, $x \mapsto \cosh \lambda x$, $x \mapsto e^{\lambda x}$, $x \mapsto (x - t)_+^\alpha$, and $x \mapsto (t - x)_+^\alpha$ belong to $\mathcal{F}^3$ for all $a \in \mathbb{R}$, $b \in \mathbb{R}$, $c \geqslant 0$, $d \geqslant 0$, $t \in \mathbb{R}$, $\alpha \geqslant 3$, and $\lambda \in \mathbb{R}$.

Note also that the classes $\mathcal{F}_+^3$, $\mathcal{F}_-^3$, and $\mathcal{F}^3$ are convex cones; that is, any linear combination with nonnegative coefficients of functions belonging to any one of these classes belongs to the same class.

**Theorem 2.10.** *Theorem 2.6 holds with $\mathcal{F}^3$ in place of $\mathcal{F}_+^3$ if the restrictions* (i) $m \geqslant m_*(p)$, (ii) $\mathsf{E}\,X_i \leqslant 0$, (iii) $\frac{b_i}{a_i} \leqslant \frac{q}{p}$, (iv) $(S_0, \dots, S_n)$ *is a supermartingale with $S_0 \leqslant 0$, and* (v) $\frac{B_{i-1}}{A_{i-1}} \leqslant \frac{q}{p}$ *in Theorem 2.6 are replaced, respectively, with the following stronger restrictions:*

(i) $m \geqslant m_*(p)$ *and $p \leqslant \frac{1}{2}$,*

(ii) $\mathsf{E}\,X_i = 0$,

(iii) $\max\left(\dfrac{b_i}{a_i}, \dfrac{a_i}{b_i}\right) \leqslant \dfrac{q}{p}$,

(iv) $(S_0, \dots, S_n)$ *is a martingale with $S_0 = 0$, and*

(v) $\max\left(\dfrac{B_{i-1}}{A_{i-1}}, \dfrac{A_{i-1}}{B_{i-1}}\right) \leqslant \dfrac{q}{p}$.

### 3. Applications

#### 3.1. Bounds on even richer classes of generalized moments (including tail probabilities), maximal inequalities, and some further extensions

Using Theorem 2.5 and Remark 1.3 (and also recalling definition (2.1) and Proposition 1.1), one immediately obtains the following corollary, which may be compared with Corollary 2.2 in [32].

**Corollary 3.1.** *Suppose that conditions* (1.15), (2.8), (2.9), *and* (2.10) *hold. Then for all $\beta \in [0, 3]$, all $f \in \mathcal{H}_+^\beta$, and all $n = 1, 2, \ldots$*

$$\mathsf{E} f(S_n) \leqslant c_{3,\beta} \, \mathsf{E} f(T_n), \tag{3.1}$$

*where*

$$T_n := s^{(m)} \cdot (\mathrm{BS}_1 + \cdots + \mathrm{BS}_n).$$

*In particular, for all real $x$,*

$$\mathsf{P}(S_n \geqslant x) \leqslant \inf_{f \in \mathcal{H}_+^3} \frac{\mathsf{E} f(T_n)}{f(x)} \tag{3.2}$$

$$= \inf_{t \in (-\infty, x)} \frac{\mathsf{E}(T_n - t)_+^3}{(x - t)^3} \tag{3.3}$$

$$\leqslant \min \left( c_{3,0} \; \mathsf{P}^{\mathsf{LC}}(T_n \geqslant x), \inf_{\lambda > 0} e^{-\lambda x} \, \mathsf{E} \, e^{\lambda T_n} \right), \tag{3.4}$$

$$= \min \left( c_{3,0} \; \mathsf{P}^{\mathsf{LC}}(T_n \geqslant x), e^{-nH} \right), \tag{3.5}$$

*where $x \mapsto \mathsf{P}^{\mathsf{LC}}(T_n \geqslant x)$ is the least log-concave majorant of the function $x \mapsto \mathsf{P}(T_n \geqslant x)$ on $\mathbb{R}$; $H := (p + y) \ln \frac{p+y}{p} + (q - y) \ln \frac{q-y}{q}$ if $0 \leqslant y := \frac{x}{n} \frac{\sqrt{pq}}{s^{(m)}} < q$, $H := -\ln p$ if $y = q$, $H := \infty$ if $y > q$, and $H := 0$ if $y < 0$.*

Note that $\mathsf{P}^{\mathsf{LC}}(T_n \geqslant x) = \mathsf{P}(T_n \geqslant x)$ for all $x$ in the lattice

$$L := \{nb + kh \colon k \in \mathbb{Z}\}$$

generated by the support of the distribution of $T_n$, where $b := s^{(m)} \sqrt{\frac{q}{p}}$ and $h := s^{(m)} / \sqrt{pq}$.

The bound $e^{-nH}$ in (3.5) is largely due to Hoeffding [18].

Using also results of [31], one has the following.

**Corollary 3.2.** *Under the conditions of Corollary 3.1,*

$$\mathsf{P}(S_n \geqslant x) \leqslant c_{3,0} \; \mathsf{P}^{\mathsf{Lin},\mathsf{LC}}(T_n \geqslant x + \tfrac{h}{2}) \quad \forall x \in \mathbb{R}, \tag{3.6}$$

*where $x \mapsto \mathsf{P}^{\mathsf{Lin},\mathsf{LC}}(T_n \geqslant x)$ is the least log-concave majorant of the linear interpolation of the tail function $x \mapsto \mathsf{P}(T_n \geqslant x)$ over the lattice $L$.*

The upper bound in (3.6) usually works better than that in (3.5) in statistical practice. An explicit formula for $\mathsf{P}^{\mathsf{Lin},\mathsf{LC}}(T_n \geqslant x + \frac{h}{2})$ is given in [31].

**Corollary 3.3.** *In view of results of [32], one can replace $S_n$ in the left-hand side of inequalities (3.2) and (3.6) by*

$$M_n := \max_{0 \leqslant k \leqslant n} S_k.$$

*Similarly, inequality (3.1) holds for all $\beta \in (0,3)$ with $M_n$ in place of $S_n$ if $c_{3,\beta}$ is replaced there by* $\dfrac{k_{1;3,\beta}}{k_{3,\beta}} c_{3,\beta}$*, where, for $\beta \in (0,\alpha)$,*

$$k_{1;\alpha,\beta} := \sup_{\sigma>0} \sigma^{-\beta(\alpha-1)} \left( \int_0^\sigma \frac{\beta s^{\beta-1}\, ds}{1+s} \right)^\alpha \quad and \quad k_{\alpha,\beta} := \frac{\beta^\beta(\alpha-\beta)^{\alpha-\beta}}{\alpha^\alpha}.$$

**Corollary 3.4.** *The set of conditions (1.15), (2.8), and (2.9) in Corollaries 3.1, 3.2, and 3.3 can be replaced by the set of conditions (1.27), (1.28), and*

$$\frac{b_i^2}{c_i^2} \leqslant \frac{q}{p} \quad \forall i = 1, \dots, n. \tag{3.7}$$

In fact, the two sets of conditions mentioned in Corollary 3.4 are equivalent to each other in a certain sense; see e.g. Remark 2.4 in [32] and the proof of Theorem 2.3 therein.

Note that the special case of Corollary 3.4 with $p = \frac{1}{2}$ (so that one may take $m = 1$) contains, among other things, Theorem 1.3 of [4], which states that, if $(S_i)$ is a martingale with $S_0 = 0$ satisfying conditions (1.27) with

$$b_i = c_i \quad \forall i = 1, \dots, n$$

and (1.28), then $\forall x \in \mathbb{R}$

$$\mathsf{P}(S_n \geqslant x) \leqslant c_{3,0}\, \mathsf{P}^{\mathsf{LC}}(T_n \geqslant x) \quad \text{with} \quad T_n = s^{(1)} \cdot (\varepsilon_1 + \cdots + \varepsilon_n); \tag{3.8}$$

by the central limit theorem, this inequality implies

$$\mathsf{P}(S_n \geqslant x) \leqslant c_{3,0}\, \mathsf{P}(s^{(1)}\sqrt{n}Z \geqslant x). \tag{3.9}$$

Obviously, inequalities (3.9) and (3.8) are extensions of (1.5) and (1.14), respectively. A version of inequality (3.9), with the larger constant factor $1/\mathsf{P}(Z > \sqrt{3}) = 24.01\ldots$ in place of $c_{3,0} = 2e^3/9 = 4.46\ldots$, appeared earlier in [3]. The improvement in the constant factor achieved in [4], as compared with [3], is due to replacing the direct method used in the earlier paper with the method based on Theorem 1.2 and Remark 1.3, which allows one to extract optimal tail comparison inequalities from comparison of generalized moments.

More generally, the generalized moments $\mathsf{E}\, f(T_n)$ in the above upper bounds, where $T_n := s^{(m)} \cdot (\mathbb{BS}_1 + \cdots + \mathbb{BS}_n)$, can be replaced by $\mathsf{E}\, f(s^{(1)}\sqrt{n}Z)$ provided that $p \geqslant \frac{1}{2}$.

That $(S_0, S_1, \dots)$ is allowed to be a supermartingale (rather than only a martingale) makes it convenient to use the simple but powerful truncation tool; cf. the discussion at the end of Section 2 in [32].

*Remark* 3.5. In the above results, the exact lower bound $m_*(p)$ on $m$ can be replaced by a substantially smaller (for $p \in (0, \frac{1}{2})$) exact lower bound $m_{\mathsf{exp}}(p)$ on $m$ if the class $\mathcal{F}_+^3$ is replaced by the substantially smaller class

$$\mathcal{F}_+^\infty := \{f \in \mathcal{C}^\infty \colon f^{(j)} \geqslant 0 \; \forall j \in \{0, 1, \dots\}\}$$

of completely monotone functions, where $\mathcal{C}^\infty$ is the class of all infinitely differentiable functions $f \colon \mathbb{R} \to \mathbb{R}$ with derivatives $f^{(0)} := f, f^{(1)} := f', \dots$. By Bernstein's theorem on completely monotone functions (see, e.g., [10] or [26]),

$$\mathcal{F}_+^\infty = \{f \colon f(x) = \int_{[0,\infty)} e^{kx} \mu(dk) \; \forall x \in \mathbb{R}\},$$

where $\mu \geqslant 0$ is a Borel measure such that the integral $\int_{[0,\infty)} e^{kx} \mu(dk)$ is finite $\forall x \in \mathbb{R}$. On the other hand, by (1.4) and Proposition 1.1,

$$\mathcal{F}_+^\infty = \{h + c \colon h \in \bigcap_{\alpha > 0} \mathcal{H}_+^\alpha, \; c \geqslant 0\}.$$

By Remark 2.3, the exact lower bound $m_{\mathsf{exp}}(p)$ cannot be less than 1, so that $m_{\mathsf{exp}}(p) = m_*(p) = 1$ for all $p \in [\frac{1}{2}, 1)$. However, for $p \in (0, \frac{1}{2})$, $m_{\mathsf{exp}}(p)$ is substantially smaller than $m_*(p)$. In particular, when $p \downarrow 0$, one has $m_{\mathsf{exp}}(p) \sim \frac{1}{2} \ln \frac{1}{p}$, while $m_*(p) \sim \frac{1}{2\sqrt{p}}$. For $p \in (0, \frac{1}{2})$, the exact lower bound $m_{\mathsf{exp}}(p)$ can be described by parametric equations

$$
\begin{cases}
m_{\mathsf{exp}}(p) = \tilde{m}(k) := \dfrac{(e^k + 1)k}{2(e^k - 1)}, \\[2mm]
\quad p = \tilde{p}(k) := \dfrac{e^k - 1 - k}{(e^k - 1)(1 + k + (k-1)e^k)},
\end{cases}
\tag{3.10}
$$

with $k \in (0, \infty)$ as the parameter. One can see that $\tilde{p}(k)$ decreases from $\frac{1}{2}$ to 0 as $k$ increases from 0 to $\infty$. An explicit upper bound on $m_{\mathsf{exp}}(p)$ is given by the inequality

$$m_{\mathsf{exp}}(p) < m_{\mathsf{exp, up}}(p) := \frac{1 - 2p - \ln(2p)}{2(1 - 2p)} \quad \forall p \in (0, \tfrac{1}{2}), \tag{3.11}$$

so that one has $m \geqslant m_{\mathsf{exp}}(p)$ provided that $m \geqslant m_{\mathsf{exp, up}}(p)$. This simple upper bound on $m_{\mathsf{exp}}(p)$ is rather good: $m_{\mathsf{exp, up}}(p) \sim m_{\mathsf{exp}}(p) \sim \frac{1}{2} \ln \frac{1}{p}$ as $p \downarrow 0$, $m_{\mathsf{exp, up}}(p) \sim m_{\mathsf{exp}}(p) \sim 1$ as $p \uparrow \frac{1}{2}$, and $\max\limits_{0 < p < 1/2} \left( m_{\mathsf{exp, up}}(p) / m_{\mathsf{exp}}(p) \right) = 1.32 \dots$, the maximum attained at $p = 0.019 \dots$.

On the other hand, the equation $\tilde{p}(k) = p$ has a unique solution $k =: \tilde{k}(p)$ in $(0, \infty)$ for every $p \in (0, \frac{1}{2})$. Moreover, the Newton iterative scheme $k_{j+1} := k_j - F(k_j)/F'(k_j)$ for $j = 0, 1, \dots$, where $F(k) := F(p, k) := e^k - 1 - k + (1 + k - 2e^k + e^{2k}(1 - k))p$, converges to $\tilde{k}(p)$ for every $p \in (0, \frac{1}{2})$ and every initial approximation $k_0 > 3(\frac{1}{\sqrt{2p}} - 1)$. Thus, one has $m_{\mathsf{exp, up}}(p) = \tilde{m}(\tilde{k}(p))$ for all $p \in (0, \frac{1}{2})$. The following table illustrates the fact that $m_{\mathsf{exp}}(p)$ is substantially smaller than $m_*(p)$ for small $p$:

| $p$ | $10^{-1}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ |
|---|---|---|---|---|
| $m_*(p)$ | 1.75 | 5.06... | 15.83... | 50.00... |
| $m_{\mathsf{exp}}(p)$ | 1.21... | 1.86... | 2.72... | 3.68... |

In particular, it follows that if $p$ is small then the second upper bound in (3.4)-(3.5), $e^{-nH}$, holds for significantly smaller values of $m$ than the first upper bound in (3.4)-(3.5), $c_{3,0}\,\mathsf{P}^{\mathsf{LC}}(T_n \geqslant x)$, does. Thus, the exponential upper bound, $e^{-nH}$, may turn out to be smaller than the generally more precise upper bound, $c_{3,0}\,\mathsf{P}^{\mathsf{LC}}(T_n \geqslant x)$, even for large values of $x$ if the $X_i$'s differ very much from one another in distribution. Details on this remark, Remark 3.5, will be presented elsewhere.

*Remark* 3.6. The stated results for the asymmetric case have "symmetric" counterparts, in which the standardized Bernoulli distribution $\mathsf{BS}(p) = p\,\delta_{\sqrt{q/p}} + q\,\delta_{-\sqrt{p/q}}$ is replaced by the standardized symmetric three-point distribution $\mathsf{ST}(p) := \frac{p}{2}\,\delta_{\sqrt{1/p}} + (1-p)\,\delta_0 + \frac{p}{2}\,\delta_{-\sqrt{1/p}}$, where $\delta_a$ is the Dirac probability measure concentrated at one point $a$ and $p \in (0,1]$. $\bigl($At that, the class $\mathcal{F}_+^3$ may be replaced by the larger, "two-tail" version $\mathcal{F}^3$. Note also that, since the distribution $\mathsf{BS}(p)$ is symmetric, one can assume without loss of generality that all generalized moment functions $f$ involved are even; indeed, for a symmetric r.v. $X$, one has $\mathsf{E}\,f(X) = \mathsf{E}\,g(X)$, where $g(x) := \frac{1}{2}\,(f(x) + f(-x))$.$\bigr)$ Then the exact lower bound $m_*(p)$ on $m$ gets replaced by another exact lower bound, whose general expression is however more complicated than expression (2.3) for $m_*(p)$. In fact, the exact lower bound on $m$ (denoted here by $m_{\mathsf{st},\,\mathsf{Schur}}(p)$) for the symmetric-case analogues of the Schur-concavity statements in Theorem 2.6 turns out to be strictly greater for some values of $p$ than the exact lower bound on $m$ (denoted here by $m_{\mathsf{st}}(p)$) for the symmetric-case analogues of the above "asymmetric" results other than Schur-concavity.

A simple upper bound on $m_{\mathsf{st},\,\mathsf{Schur}}(p)$ and hence on $m_{\mathsf{st}}(p)$ for all $p \in (0,\frac{1}{2}]$ is given by inequality

$$m_{\mathsf{st}}(p) \leqslant m_{\mathsf{st},\,\mathsf{Schur}}(p) \leqslant m_{\mathsf{st},\,\mathsf{high}}(p) := m_*(r) = \frac{5 - 3\sqrt{1-2p} - 2p}{4(\sqrt{\frac{p}{2}} + 1 - \sqrt{1-2p} - p)} \quad (3.12)$$

where the function $m_*$ is defined by (2.3) and $r := (1 - \sqrt{1-2p})/2$ is the root in $(0,\frac{1}{2}]$ of equation $2r(1-r) = p$. This follows because the convolution of the standardized Bernoulli distributions $\mathsf{BS}(r)$ and $\mathsf{BS}(1-r)$ is a symmetric three-point distribution, which ascribes probability $1-p$ to $\{0\}$ and therefore can be obtained from the standardized symmetric three-point distribution $\mathsf{ST}(p)$ by simple re-scaling. One can see that $m_{\mathsf{st},\,\mathsf{high}}(p) \sim \frac{1}{\sqrt{2p}}$ as $p \downarrow 0$ and $m_{\mathsf{st},\,\mathsf{high}}(\frac{1}{2}) = 1$; let also $m_{\mathsf{st},\,\mathsf{high}}(p) := 1$ for all $p \in (\frac{1}{2},1)$.

A remarkable and not so difficult to prove fact is that

$$m_{\mathsf{st}}(p) = 1 \quad \forall p \in [p_*,1], \quad \text{where} \quad p_* := \sqrt{2} - 1 = 0.4142\ldots. \quad (3.13)$$

More generally, it appears that

$$m_{\mathsf{st}}(p) = m_1(p)\boldsymbol{I}\{0 < p < p_{0,1}\} + m_0(p)\boldsymbol{I}\{p_{0,1} \leqslant p < p_*\} + \boldsymbol{I}\{p_* \leqslant p < 1\},$$

where $p_{0,1} = 0.3878\ldots$, $m_1(p) := \frac{1}{2}\sqrt{\frac{2-p}{p}}$, and $m_0(p) := \frac{1}{2\log_2 z}$, where $z$ is (for each $p \in (0, p_*)$) the only root in the interval $(0, \sqrt{2})$ of polynomial

$$(9p^2 - 24p + 16)z^6 + (-36p^2 + 120p - 96)z^5 + (36p^2 - 216p + 240)z^4$$
$$+ (-20p^3 + 60p^2 + 120p - 320)z^3 + (72p^3 - 168p^2 + 96p + 240)z^2$$
$$+ (-96p^3 + 144p^2 - 144p - 96)z + 4p^4 + 40p^3 - 44p^2 + 48p + 16.$$

(Here, as usual, $\boldsymbol{I}\{\mathcal{A}\}$ denotes the indicator of an assertion $\mathcal{A}$.)

On the other hand, one can show that a lower bound on $m_{\mathsf{st}}(p)$ and hence on $m_{\mathsf{st,Schur}}(p)$ is given by the inequality

$$m_{\mathsf{st,Schur}}(p) \geqslant m_{\mathsf{st}}(p) \geqslant m_{\mathsf{st,low}}(p) := \max(1, m_1(p), m_{\mathsf{low}}(p))$$
$$= m_1(p)\boldsymbol{I}\{0 < p < p_{\mathsf{low},1}\} + m_{\mathsf{low}}(p)\boldsymbol{I}\{p_{\mathsf{low},1} \leqslant p < p_*\} + \boldsymbol{I}\{p_* \leqslant p < 1\}, \tag{3.14}$$

where $p_{\mathsf{low},1} = 0.3889\ldots$ and $m_{\mathsf{low}}(p) := \frac{3}{2(1+\log_2(1+p))}$. It appears that this lower bound, $m_{\mathsf{st,low}}(p)$ is very close to $m_{\mathsf{st}}(p)$ and differs from the latter only for $p$ in the rather narrow interval $(p_{0,1}, p_*) \approx (0.3878, 0.4142)$, and at most by $m_{\mathsf{st}}(p_{\mathsf{low},1}) - m_{\mathsf{st,low}}(p_{\mathsf{low},1}) = 0.0008598\ldots$. It follows from (3.14) that the lower bound $p_* = \sqrt{2} - 1$ on $p$ in (3.13) is exact; that is, $m_{\mathsf{st}}(p) > 1 \ \forall p \in (0, p_*)$. Also, the four bounds, $m_{\mathsf{st,low}}(p) \leqslant m_{\mathsf{st}}(p) \leqslant m_{\mathsf{st,Schur}}(p) \leqslant m_{\mathsf{st,high}}(p)$, all equal 1 for all $p \in [\frac{1}{2}, 1)$ and are asymptotic to $1/\sqrt{2p}$ as $p \downarrow 0$; moreover, the difference between any two of these four bounds goes to 0 as $p \downarrow 0$, since it is easy to see that $m_{\mathsf{st,high}}(p) - m_{\mathsf{st,low}}(p) \to 0$ as $p \downarrow 0$.

In this "symmetric" setting, the place of supermartingales gets taken by conditionally symmetric martingales (that is, martingales $(S_i)$ with $S_0 = 0$ a.s. and conditionally symmetric differences $X_i$), and "bounded-asymmetry" condition (3.7) in Corollary 3.4 gets replaced by "bounded-kurtosis" condition

$$\frac{b_i^2}{c_i^2} \leqslant \frac{1}{p} \quad \forall i = 1, \ldots, n. \tag{3.15}$$

(Note that the kurtosis of a standardized symmetric three-point r.v. $\mathsf{ST} \sim \mathsf{ST}(p)$ is $\frac{1}{p}$.) Thus, one now has another – "symmetric" – version of inequality (3.8), which immediately implies a striking "symmetric" version of inequality (3.9); namely, in view of (3.13) and (3.15), inequality (3.9) holds for all conditionally symmetric martingales if the condition $b_i = c_i \ \forall i$ of [4, Theorem 1.3] is replaced by the much less restrictive condition

$$b_i^2 \leqslant \frac{c_i^2}{\sqrt{2} - 1} \approx 2.4142 \, c_i^2 \quad \forall i = 1, \ldots, n.$$

In particular, one has the following generalizations of (1.1) and (1.5), respectively:

$$\mathsf{E} \, f\left(a_1 \mathsf{ST}_1 + \cdots + a_n \mathsf{ST}_n\right) \leqslant \mathsf{E} \, f(Z) \quad \forall f \in \mathcal{F}^3 \tag{3.16}$$

and

$$\mathsf{P}\left(a_1 \mathsf{ST}_1 + \cdots + a_n \mathsf{ST}_n \geqslant x\right) \leqslant \frac{2e^3}{9} \, \mathsf{P}(Z \geqslant x) \quad \forall x \in \mathbb{R}, \tag{3.17}$$

where $\mathrm{ST}_i \overset{\text{i.i.d.}}{\sim} \mathsf{ST}(p)$ and $p \in [\sqrt{2}-1, 1]$; note that $\mathsf{ST}(p)$ coincides with the distribution of a Rademacher r.v. if $p = 1$.

Inequalities for the much simpler case when $b_i = b \;\forall i$ were obtained in [8, 5] (in the "exponential" case) and in [6].

One also has multi-dimensional analogues of inequalities (3.16) and (3.17). Namely, for all $p \in [\sqrt{2}-1, 1]$ one has the following generalizations of inequalities (1.12) and (1.13):

$$\mathsf{E}\, f\left(\|\mathrm{ST}_1 \mathbf{x}_1 + \cdots + \mathrm{ST}_n \mathbf{x}_n\|\right) \leqslant \mathsf{E}\, f(|Z|) \quad \text{for all even } f \text{ in } \mathcal{F}^3 \qquad (3.18)$$

and

$$\mathsf{P}\left(\|\mathrm{ST}_1 \mathbf{x}_1 + \cdots + \mathrm{ST}_n \mathbf{x}_n\| \geqslant x\right) \leqslant \frac{2e^3}{9}\, \mathsf{P}(|Z| \geqslant x) \quad \forall x \in \mathbb{R}, \qquad (3.19)$$

with the Rademacher r.v.'s $\varepsilon_1, \dots, \varepsilon_n$ replaced by $\mathrm{ST}_1, \dots, \mathrm{ST}_n$.

Condition (3.15) may be interpreted as a condition of boundedness of the kurtoses of the $X_i$'s; cf. bounded-asymmetry conditions (2.6) and (2.9). Note that the usual interpretation of the term "kurtosis" (as well as the term itself, meaning "peakedness") is not quite adequate, especially in contexts of large deviations. Indeed, without loss of generality, let us assume that a r.v. $X$ is standardized, that is, zero-mean and unit-variance. Then the kurtosis of $X$ is $\mathrm{kurt}(X) = \mathsf{E}\, X^4 = 1 + \mathsf{E}(X^2-1)^2 = 1 + \mathsf{Var}(X^2)$, which clearly is the measure of relative spread of the values of the nonnegative r.v. $X^2$ about its mean 1. Thus, $\mathrm{kurt}(X)$ is large if and and only if at least one of the two tails, left or right, of the distribution of $X^2$ on $[0, \infty)$ is heavy. These two tails can be measured, respectively, by the "negative" and "positive" parts of $\mathrm{kurt}(X) - 1$, namely $\mathrm{kurt}_-(X) := \mathsf{E}(X^2-1)^2 \boldsymbol{I}\{X^2 < 1\}$ and $\mathrm{kurt}_+(X) := \mathsf{E}(X^2-1)^2 \boldsymbol{I}\{X^2 > 1\}$. If $\mathrm{kurt}_-(X)$ is large, then the distribution of $X$ is peaked at 0; if $\mathrm{kurt}_+(X)$ is large, then at least one tail of the distribution of $X$ is heavy. Of these two reasons for the kurtosis of $X$ to be large, the heaviness of the tails of the distribution of $X$ seems more important than the peakedness, at least where large deviations are involved. Also, note that the kurtosis of $X$ can be infinite only if the tails of the distribution of $X$ are heavy enough. As was noted, the kurtosis of a standardized symmetric three-point r.v. $\mathrm{ST} \sim \mathsf{ST}(p)$ is $\frac{1}{p}$, which is decreasing in $p$; at that, both the negative part $\mathrm{kurt}_-(\mathrm{ST}) = 1 - p$ and the positive part $\mathrm{kurt}_+(\mathrm{ST}) = 4 \sinh^2 \ln \frac{1}{\sqrt{p}}$ are decreasing in $p$. More generally, it is not difficult to show that for any functions $f \in \mathcal{F}^3$ (except for $f$ of the form $f(x) \equiv a + bx^2$ for some nonnegative real constants $a$ and $b$) the generalized moments $\mathsf{E}\, f(\mathrm{ST})$ and hence $\mathsf{E}\, f(c_1 \mathrm{ST}_1 + \cdots + c_n \mathrm{ST}_n)$ are strictly decreasing in $p \in (0, 1]$, where $\mathrm{ST}, \mathrm{ST}_i \overset{\text{i.i.d.}}{\sim} \mathsf{ST}(p)$ and the $c_i$'s are nonzero real constants. Thus, inequality (3.16) for $p \in [\sqrt{2}-1, 1)$ not only generalizes but also strengthens inequality (1.1). Similarly, inequality (3.17) is stronger than inequality (1.5) for $p \in [\sqrt{2}-1, 1)$ and large enough $x$.

It may also be instructive to compare inequality (3.17) with available asymptotic results. For instance, it follows from [16] or [25] that for *every* $p \in (0, 1)$

$$\mathsf{P}\,(\mathrm{ST}_1 + \cdots + \mathrm{ST}_n \geqslant x) \sim \mathsf{P}(Z \geqslant x/\sqrt{n}) \tag{3.20}$$

provided that $n \to \infty$ and $x = o(n)$. The advantages of this asymptotics over inequality (3.17) are (i) that (3.20) holds for every $p \in (0,1)$ and not just for $p \in [\sqrt{2}-1, 1]$ and (ii) there is no "extra" constant factor (such as $2e^3/9$) in the right-hand side of (3.20). However, these advantages seem to be counterbalanced by the following: (i) the asymptotic relation (3.20), without an upper bound on the rate of convergence, is impossible to use in statistical practice when one needs to be certain that the left-hand side of (3.20) does not exceed a prescribed level; (ii) the asymptotics is valid only in the zone $x = o(n)$, and this zone is defined only qualitatively; (iii) the summands $\mathrm{ST}_1, \ldots, \mathrm{ST}_n$ in the left-hand side of (3.20) are identically distributed; if coefficients $a_1, \ldots, a_n$ are allowed in (3.20) (as in (3.17)), they cannot differ too much from another, and also then the zone $x = o(n)$ must be narrowed; (iv) there is no "extra" constant factor (such as $2e^3/9$) in the right-hand side of the generalized-moment comparison inequality (3.16); (v) one can use the more precise upper bounds of the form given in (1.9) or (1.33) instead of (3.17).

Details on this long remark, Remark 3.6, will be presented elsewhere.

*Remark* 3.7. Directions of further research indicated in Remarks 3.5 and 3.6 can be combined. That is, one can at once replace (i) the class of generalized moment functions $\mathcal{F}_+^3$ by the much narrower class $\mathcal{F}_+^\infty$ (the latter being in effect the same as the class of all nondecreasing exponential functions) and (ii) the standardized Bernoulli distribution by the standardized symmetric three-point distribution. At that, one has another exact lower bound, say $m_{\mathsf{st},\exp}(p)$, on $m$. Obviously, $m_{\mathsf{st},\exp}(p) \leqslant m_{\mathsf{st}}(p)$ for all $p \in (0,1]$. In particular, $m_{\mathsf{st},\exp}(p) = m_{\mathsf{st}}(p) = 1$ for all $p \in [\sqrt{2}-1, 1]$. Also, similarly to (3.12) and in view of (3.11), one has $m_{\mathsf{st},\exp}(p) \leqslant m_{\exp}(r) \leqslant m_{\exp,\mathsf{up}}(r)$ for all $p \in (0, \sqrt{2}-1)$ (and even for all $p \in (0, \frac{1}{2}])$, where as before $r = (1 - \sqrt{1-2p})/2$ is the root in $(0, \frac{1}{2}]$ of equation $2r(1-r) = p$. While, as noted, $m_{\mathsf{st},\exp}(p) = 1$ for all $p \in [\sqrt{2}-1, 1]$, one has $m_{\mathsf{st},\exp}(p) > 1$ for all $p \in (0, \frac{1}{3})$. Indeed, if $m_{\mathsf{st},\exp}(p) = 1$ for some $p \in (0,1)$, then inequality (3.16) holds for that same $p$ and all $f \in \mathcal{F}_+^\infty$. Taking now $n = 1$ and $f(x) \equiv e^{kx}$ in (3.16), and then letting $k \downarrow 0$, one has $\frac{1}{p} = \mathsf{E}\,\mathrm{ST}_1^4 \leqslant \mathsf{E}\,Z^4 = 3$, whence $p \geqslant \frac{1}{3}$.

Details on this remark, Remark 3.7, will be presented elsewhere.

### 3.2. Bounds on self-normalized sums

(Details on the results presented in this subsection will be given elsewhere.) Efron [15] considered the so-called self-normalized sum

$$V := \frac{X_1 + \cdots + X_n}{\sqrt{X_1^2 + \cdots + X_n^2}}, \tag{3.21}$$

assuming that the $X_i$'s satisfy the orthant symmetry condition: the joint distribution of $\delta_1 X_1, \ldots, \delta_n X_n$ is the same for any choice of signs $\delta_1, \ldots, \delta_n \in \{1, -1\}$,

so that, in particular, each $X_i$ is symmetric(ally distributed). It suffices that the $X_i$'s be independent and symmetrically (but not necessarily identically) distributed. On the event $\{X_1 = \cdots = X_n = 0\}$, let $V := 0$.

In Subsection 3.2 we assume throughout that the $X_i$'s are all non-degenerate: $\mathsf{P}(X_i = 0) < 1$ for all $i$.

Note that the conditional distribution of any symmetric r.v. $X$ given $|X|$ is the symmetric distribution on the (at most) two-point set $\{X, -X\}$. Therefore, under the orthant symmetry condition, the distribution of $V$ is the mixture of the distributions of the normalized Khinchin-Rademacher sums $a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$, where $a_i = X_i/(X_1^2 + \cdots + X_n^2)^{\frac{1}{2}}$, so that $a_1^2 + \cdots + a_n^2 = 1$ (except on the event $\{X_1 = \cdots = X_n = 0\}$, where $a_1 = \cdots = a_n = 0$). Hence, by (1.1) (for $f(x) \equiv e^{\lambda x}$, $\lambda \geqslant 0$) and (1.2) one has

$$\mathsf{E}\, e^{\lambda V} \leqslant \mathsf{E}\, e^{\lambda Z} \quad \forall \lambda \geqslant 0 \tag{3.22}$$

and

$$\mathsf{P}\,(V \geqslant x) \leqslant e^{-x^2/2} \quad \forall x \geqslant 0. \tag{3.23}$$

These results can be easily restated in terms of Student's statistic $T$, which is a monotonic function of $V$, as noted by Efron; namely, $T = \sqrt{\frac{n-1}{n}}\, V/\sqrt{1 - V^2/n}$. Inequalities (3.22) and (3.23) were improved in [12, 13, 27] using (1.1) (for $f \in \mathcal{H}_+^3$) and (1.5), so that one has

$$\mathsf{E}\, f(V) \leqslant \mathsf{E}\, f(Z) \quad \forall f \in \mathcal{H}_+^3 \tag{3.24}$$

and

$$\mathsf{P}\,(V \geqslant x) \leqslant \frac{2e^3}{9}\, \mathsf{P}(Z \geqslant x) \quad \forall x \in \mathbb{R}. \tag{3.25}$$

Multivariate analogues of these results, which can be expressed in terms of Hotelling's statistic in place of Student's, were also obtained in [27].

It was pointed out in [27, Theorem 2.8] that, since the normal tail decreases fast, inequality (3.25) implies that relevant quantiles of $V$ may exceed the corresponding standard normal quantiles only by a relatively small amount, so that one can use (3.25) rather efficiently to test symmetry even for non-i.i.d. observations.

Here we shall present extensions of inequalities (3.24) and (3.25) to the case when the $X_i$'s are not symmetric, as well as improvements of (3.24) and (3.25) in the "symmetric case". Asymptotics for large deviations of $V$ for i.i.d. $X_i$'s without moment conditions was obtained recently by Jing, Shao and Zhou [19]. Comments similar to the ones made at the end of Remark 3.6 apply here as well.

### 3.2.1. The "asymmetric" case

The basic idea here is to represent any zero-mean, possibly asymmetric distribution as an appropriate mixture of two-point zero-mean distributions. Let us

assume at first that a zero-mean r.v. $X$ has an everywhere strictly positive density function. Consider the truncated r.v. $X_{a,b} := X\boldsymbol{I}\{a \leqslant X \leqslant b\}$. Then, for every fixed $a \in (-\infty, 0]$, the function $b \mapsto \mathsf{E}\,X_{a,b}$ is continuous and increasing on the interval $[0, \infty)$ from $\mathsf{E}\,X_{a,0} \leqslant 0$ to $\mathsf{E}\,X_{a,\infty} > 0$. Hence, for each $a \in (-\infty, 0]$, there exists a unique value $b \in [0, \infty)$ such that $\mathsf{E}\,X_{a,b} = 0$. Similarly, for each $b \in [0, \infty)$, there exists a unique value $a \in (-\infty, 0]$ such that $\mathsf{E}\,X_{a,b} = 0$. That is, one has a one-to-one correspondence between $a \in (-\infty, 0]$ and $b \in [0, \infty)$ such that $\mathsf{E}\,X_{a,b} = 0$. Denote by $\mathsf{r} := \mathsf{r}_X$ the *reciprocating* function defined on $\mathbb{R}$ and carrying this correspondence, so that

$$\mathsf{E}\,X\boldsymbol{I}\{X \text{ is between } x \text{ and } \mathsf{r}(x)\} = 0 \quad \forall x \in \mathbb{R},$$

the function $\mathsf{r}$ is decreasing on $\mathbb{R}$ and such that $\mathsf{r}(\mathsf{r}(x)) = x \ \forall x \in \mathbb{R}$; moreover, $\mathsf{r}(0) = 0$. (Clearly, $\mathsf{r}(x) = -x$ for all real $x$ if the r.v. $X$ is symmetric.) One also has

$$\mathsf{r}(x) = x_-(G(x))\,\boldsymbol{I}\{x > 0\} + x_+(G(x))\,\boldsymbol{I}\{x < 0\}, \tag{3.26}$$

where $x_\pm(h)$ stand for the positive and negative roots $x$ of the equation $G(x) = h$ and, in turn,

$$G(x) := \mathsf{E}\,|X|\boldsymbol{I}\{|X| \leqslant |x|, \ \mathrm{sign}\,X = \mathrm{sign}\,x\}. \tag{3.27}$$

Thus, the set $\{\,\{x, \mathsf{r}(x)\} \colon x \in \mathbb{R}\,\}$ of (at-most-)two-point sets constitutes a partition of $\mathbb{R}$. Moreover, the two-point set $\{x, \mathsf{r}(x)\}$ is uniquely determined by the distance $|x - \mathsf{r}(x)|$ between the two points, as well as by the product $|x|\,|\mathsf{r}(x)|$. Now one can see that the conditional distribution of the zero-mean r.v. $X$ given $W := |X - \mathsf{r}(X)|$ (or, equivalently, $Y := |X\,\mathsf{r}(X)|$) is the uniquely determined zero-mean distribution on the two-point set $\{X, \mathsf{r}(X)\}$. Thus, the distribution of the zero-mean r.v. $X$ with an everywhere positive density is represented as a mixture of two-point zero-mean distributions. This mixture is given rather explicitly, provided that the distribution of r.v. $X$ is known.

Thus, one has generalized versions of the self-normalized sum (3.21), which require – instead of the symmetry of independent r.v.'s $X_i$ – only that the $X_i$'s be zero-mean:

$$V_W := \frac{X_1 + \cdots + X_n}{\frac{1}{2}\sqrt{W_1^2 + \cdots + W_n^2}} \quad \text{and} \quad V_{Y,m} := \frac{X_1 + \cdots + X_n}{(Y_1^m + \cdots + Y_n^m)^{\frac{1}{2m}}}, \tag{3.28}$$

where $m \geqslant 1$,

$$W_i := |X_i - \mathsf{r}_i(X_i)|, \quad \text{and} \quad Y_i := |X_i\,\mathsf{r}_i(X_i)|,$$

and the reciprocating function $\mathsf{r}_i := \mathsf{r}_{X_i}$ is constructed as above, based on the distribution of $X_i$, for each $i$, so that the reciprocating functions $\mathsf{r}_i$ may be different from one another if the $X_i$'s are not identically distributed. On the event $\{X_1 = \cdots = X_n = 0\}$ (which is the same as either one of events $\{W_1 = \cdots = W_n = 0\}$ and $\{Y_1 = \cdots = Y_n = 0\}$), let $V_W := 0$ and $V_{Y,m} := 0$. Note that $V_W = V_{Y,1} = V$ when the $X_i$'s are symmetric. Logan *et al* [22] and

Shao [35] obtained limit theorems for the "symmetric" version of $V_{Y,m}$ (with the reciprocating function $\mathsf{r}(x) \equiv -x$), whereas the $X_i$'s did not need to be symmetric.

These constructions can be extended to the general case of any zero-mean r.v. $X$, absolutely continuous or not. Here, one can use randomization (by means of a r.v. uniformly distributed in interval $(0,1)$) to deal with the atoms of the distribution of r.v. $X$, and a modification of the inverse functions $x_{\pm}(h)$ to deal with the intervals on which the distribution function of $X$ and hence the function $G$ are constant. Namely, in general $\mathsf{r}(X)$ is replaced by $\mathsf{r}(X, U)$, where $U$ is a r.v. uniformly distributed in interval $(0,1)$ and independent of $X$ and, for $x \in \mathbb{R}$ and $u \in (0,1)$,

$$\mathsf{r}(x, u) := \begin{cases} x_-(G(x-) + u \cdot (G(x) - G(x-))) & \text{if } x \in [0, \infty), \\ x_+(G(x+) + u \cdot (G(x) - G(x+))) & \text{if } x \in (-\infty, 0], \end{cases}$$
$$x_+(h) := \inf\{x \in [0, \infty] \colon G(x) \geqslant h\},$$
$$x_-(h) := \sup\{x \in [-\infty, 0] \colon G(x) \geqslant h\}.$$

By conditioning on the $W_i$'s or $Y_i$'s one obtains the following corollaries.

**Corollary 3.8.** (From results of [32]:)

$$\mathsf{E}\, f(V_W) \leqslant \mathsf{E}\, f(Z) \quad \forall f \in \mathcal{H}_+^5, \quad whence \tag{3.29}$$
$$\mathsf{P}(V_W \geqslant x) \leqslant c_{5,0}\, \mathsf{P}(Z \geqslant x) \quad \forall x \in \mathbb{R}, \tag{3.30}$$

*where, in accordance with* (1.11), $c_{5,0} = 5!(e/5)^5 = 5.699\ldots$.

**Corollary 3.9.** (From Theorem 2.2 and Corollary 3.1:) *Suppose that for some* $p \in (0, 1)$ *and all* $i \in \{1, \ldots, n\}$

$$\frac{X_i}{|\mathsf{r}(X_i)|}\, \boldsymbol{I}\{X_i > 0\} \leqslant \frac{q}{p} \quad a.s. \tag{3.31}$$

*Then for all* $m \geqslant m_*(p)$

$$\mathsf{E}\, f(V_{Y,m}) \leqslant \mathsf{E}\, f(T_n) \quad \forall f \in \mathcal{F}_+^3 \quad and \tag{3.32}$$
$$\mathsf{P}(V_{Y,m} \geqslant x) \leqslant c_{3,0}\, \mathsf{P}^{\mathsf{LC}}(T_n \geqslant x) \quad \forall x \in \mathbb{R}, \tag{3.33}$$

*where* $T_n$ *and* $\mathsf{P}^{\mathsf{LC}}(T_n \geqslant x)$ *have the same meaning as in Corollary 3.1 and, in accordance with* (1.11), $c_{3,0} = 2e^3/9 = 4.4634\ldots$.

Of course, one can replace the upper bound in inequalities like (3.33) by either of the more precise (but slightly less transparent and more difficult to compute) upper bounds given in (3.3) and (3.6).

By Remark 3.5, inequality (3.32) will continue to hold for all $m \geqslant m_{\mathsf{exp}}(p)$ provided that the class $\mathcal{F}_+^3$ is replaced by the smalled class $\mathcal{F}_+^\infty$; correspondingly, for such $m$ one will have the exponential upper bound on $\mathsf{P}(V_{Y,m} \geqslant x)$ of the form $e^{-nH}$ as in (3.5).

Condition (3.31) is likely to hold when the $X_i$'s are bounded i.i.d. r.v.'s.

Note that the reciprocating function $\mathsf{r}$ depends on the (usually unknown in statistics) distribution of the underlying r.v. $X$. However, if e.g. the $X_i$ constitute an i.i.d. sample, then the function $G$ defined by (3.27) can be estimated based on the sample, so that one can estimate the reciprocating function $\mathsf{r}$. Thus, replacing $X_1 + \cdots + X_n$ in the numerators of $V_W$ and $V_{Y,m}$ by $X_1 + \cdots + X_n - n\theta$, one obtains approximate pivots to be used to construct confidence intervals or, equivalently, tests for an unknown mean $\theta$. One can also use bootstrap to estimate the distributions of such pivots.

### 3.2.2. The "symmetric" case

Here we assume that $X_1, \ldots, X_n$ are independent symmetric r.v.'s. In this case, we already have the upper bounds given by (3.24) and (3.25), which are based on (1.1) and (1.5). As was pointed out, bounds (3.16) and (3.17) improve and generalize (1.1) and (1.2). Correspondingly, the upper bounds given by (3.24) and (3.25) can be improved and generalized as follows.

Introduce another, "symmetric" modification of the standard self-normalized sum $V$ given by (3.21):

$$V_{\mathsf{symm},m,p} := \frac{\mathrm{ST}_1 X_1 + \cdots + \mathrm{ST}_n X_n}{(|X_1|^{2m} + \cdots + |X_n|^{2m})^{\frac{1}{2m}}}, \tag{3.34}$$

where $p \in (0,1)$, $m \geqslant 1$, $\mathrm{ST}_1, \ldots, \mathrm{ST}_n$ are i.i.d. $\mathsf{ST}(p)$ r.v.'s, independent also of $X_1, \ldots, X_n$. On the event $\{X_1 = \cdots = X_n = 0\}$, let $V_{\mathsf{symm},m,p} := 0$.

Thus, the distribution of $V_{\mathsf{symm},m,p}$ is the mixture of the distributions of the linear combinations of the form $a_1 \mathrm{ST}_1 + \cdots + a_n \mathrm{ST}_n$, where $a_i = X_i/(|X_1|^{2m} + \cdots + |X_n|^{2m})^{\frac{1}{2m}}$, so that $|a_1|^{2m} + \cdots + |a_n|^{2m} = 1$ (except on the event $\{X_1 = \cdots = X_n = 0\}$, where $a_1 = \cdots = a_n = 0$).

By Remark 3.6, it follows that for all $m \geqslant m_{\mathsf{st}}(p)$

$$\mathsf{E}\, f(V_{\mathsf{symm},m,p}) \leqslant \mathsf{E}\, f(n^{-1/(2m)}(\mathrm{ST}_1 + \cdots + \mathrm{ST}_n)) \quad \forall f \in \mathcal{F}^3, \quad \text{whence} \tag{3.35}$$

$$\mathsf{P}(V_{\mathsf{symm},m,p} \geqslant x) \leqslant c_{3,0}\, \mathsf{P}^{\mathsf{LC}}(n^{-1/(2m)}(\mathrm{ST}_1 + \cdots + \mathrm{ST}_n) \geqslant x) \quad \forall x \in \mathbb{R}, \tag{3.36}$$

where $\mathsf{P}^{\mathsf{LC}}$ again denotes the least log-concave majorant of the corresponding tail function. In particular (cf. (3.16) and (3.17)), for all $p \in [\sqrt{2} - 1, 1)$,

$$\mathsf{E}\, f(V_{\mathsf{symm},1,p}) \leqslant \mathsf{E}\, f(n^{-1/2}(\mathrm{ST}_1 + \cdots + \mathrm{ST}_n)) \tag{3.37}$$

$$\leqslant \mathsf{E}\, f(Z) \quad \forall f \in \mathcal{F}^3, \tag{3.38}$$

$$\mathsf{P}(V_{\mathsf{symm},1,p} \geqslant x) \leqslant c_{3,0}\, \mathsf{P}^{\mathsf{LC}}(n^{-1/2}(\mathrm{ST}_1 + \cdots + \mathrm{ST}_n) \geqslant x) \tag{3.39}$$

$$\leqslant c_{3,0}\, \mathsf{P}(Z \geqslant x) \quad \forall x \in \mathbb{R}. \tag{3.40}$$

In view of the mixture representations of the distributions of $V$ and $V_{\mathsf{symm},m,p}$ and the discussion after inequalities (3.16)–(3.19), inequalities (3.35)–(3.40) for $V_{\mathsf{symm},m,p}$ generalize/improve inequalities (3.22)–(3.25) for $V$.

The classic self-normalized sum $V$ can be obviously used, e.g., as a test statis- tic to test the symmetry of the distributions of the $X_i$'s. However, it is not seen how $V_{\mathsf{symm},m,p}$ can be used in symmetry tests, because its distribution will be always symmetric, even if the distributions of the $X_i$'s are not.

On the other hand, there are two issues with $V$:

**(i)** $V$ may have too light tails if the tails of the $X_i$'s are heavy enough, and so, the symmetry test based on such an upper bound as the one given by inequality (3.25) may turn out to be too conservative and hence lacking some power;

**(ii)** the way $V$ (as well as its modifications considered so far – $V$, $V_W$, $V_{Y,m}$, and $V_{\mathsf{symm},m,p}$) "deals" with the event when all the $X_i$'s take on the zero value certainly seems to be too conservative; note that this event may naturally occur with a nonzero probability if the distribution of $X_i$ is discrete or if the original r.v.'s $X_i$ are replaced by the corresponding truncated r.v.'s $X_i \boldsymbol{I}\{|X_i| < b_i\}$ for some $b_i > 0$ (which may be done to increase the power of the test). However, note that usually the probabilities $p_i := \mathsf{P}(X_i \neq 0)$ will be close to (even if less than) 1.

To try to resolve these two issues with $V$, we shall suggest yet another mod- ification of it. Each of the other self-normalized sums introduced above – $V_W$, $V_{Y,m}$, and $V_{\mathsf{symm},m,p}$ – can be modified in the same manner.

Let us accompany any r.v. $X$ with a r.v. of the form

$$\hat{X} := X + \tilde{X}\,\boldsymbol{I}\{X = 0\}\,, \tag{3.41}$$

where $\tilde{X}$ is a r.v. which is independent of $X$ and whose distribution coincides with the conditional distribution of $X$ given that $X \neq 0$; thus, $\hat{X} = X$ if $X \neq 0$ and $\hat{X} = \tilde{X}$ if $X = 0$, so that $\mathsf{P}(\hat{X} \neq 0) = 1$. It is not hard to see that, if $X$ is symmetric, then the conditional distribution of $X$ given $|\hat{X}|$ coincides with the symmetric distribution $\frac{p}{2}\delta_{\hat{X}} + (1-p)\delta_0 + \frac{p}{2}\delta_{-\hat{X}}$ on the three-point set $\{\hat{X}, 0, -\hat{X}\}$, where $p := \mathsf{P}(X \neq 0)$:

$$\mathcal{L}(X|\hat{X}) = \tfrac{p}{2}\delta_{\hat{X}} + (1-p)\delta_0 + \tfrac{p}{2}\delta_{-\hat{X}}. \tag{3.42}$$

Moreover, the distribution of $\hat{X}$ coincides (just as that of $\tilde{X}$ does) with the conditional distribution of $X$ given that $X \neq 0$. Indeed, $\hat{X}$ is equal in distribution to the first nonzero member of an infinite random sequence $(X^{(1)}, X^{(2)}, \ldots)$, if $X, X^{(1)}, X^{(2)}, \ldots$ are i.i.d. r.v.'s; therefore, one may assume that $\hat{X}$ equals (as a r.v., and not just in distribution) to the first nonzero member of the sequence $(X, X^{(1)}, X^{(2)}, \ldots)$. Thus, roughly speaking, to get $\hat{X}$, one samples from the distribution of $X$ till getting a nonzero replica of $X$. Note also that for any even function $g$ such that $g(0) = 0$ one has

$$\mathsf{Var}\, g(\hat{X}) - \mathsf{Var}\, g(X) = \tfrac{1-p}{p}\left(\mathsf{Var}\, g(X) - \tfrac{1}{p}(\mathsf{E}\, g(X))^2\right). \tag{3.43}$$

It follows from (3.42) that the distribution of yet another modification of the

self-normalized sum $V$,

$$\hat{V}_{\mathsf{symm},m,p} := \frac{X_1 + \cdots + X_n}{\sqrt{p}\,(|\hat{X}_1|^{2m} + \cdots + |\hat{X}_n|^{2m})^{\frac{1}{2m}}}, \qquad (3.44)$$

is the mixture of the distributions of the linear combinations of the form $a_1\mathrm{ST}_1(p_1) + \cdots + a_n\mathrm{ST}_n(p_n)$, where the $\mathrm{ST}_i(p_i)$'s are independent r.v.'s such that $\mathrm{ST}_i(p_i) \sim \mathsf{ST}(p_i)$ for all $i$ and $a_i = \hat{X}_i/(|\hat{X}_1|^{2m} + \cdots + |\hat{X}_n|^{2m})^{\frac{1}{2m}}$, so that $|a_1|^{2m} + \cdots + |a_n|^{2m} = 1$ a.s.; here, $p \in (0,1)$,

$$p \geqslant p_i := \mathsf{P}(X_i \neq 0) \quad \forall i,$$

and each $\hat{X}_i$ is produced based on $X_i$ according to formula (3.41), where the $\tilde{X}_i$'s are independent of one another and of the $X_i$'s.

Recall that, for each $i$, one has $\mathsf{P}(\hat{X}_i = 0) = 0$ and the distribution of $\hat{X}_i$ coincides with the conditional distribution of $X_i$ given that $X_i \neq 0$.

It follows that

$$\mathsf{E}\,f(\hat{V}_{\mathsf{symm},m,p}) \leqslant \mathsf{E}\,f(n^{-1/(2m)}(\mathrm{ST}_1 + \cdots + \mathrm{ST}_n)) \quad \forall f \in \mathcal{F}^3, \qquad (3.45)$$

whence

$$\mathsf{P}(\hat{V}_{\mathsf{symm},m,p} \geqslant x) \leqslant c_{3,0}\,\mathsf{P}^{\mathsf{LC}}(n^{-1/(2m)}(\mathrm{ST}_1 + \cdots + \mathrm{ST}_n) \geqslant x) \quad \forall x \in \mathbb{R}, \quad (3.46)$$

provided that

$$m \geqslant m_{\mathsf{st}}(p);$$

in particular, if $p \geqslant \sqrt{2} - 1$ (which will typically be the case),

$$\mathsf{E}\,f(\hat{V}_{\mathsf{symm},1,p}) \leqslant \mathsf{E}\,f(n^{-1/2}(\mathrm{ST}_1 + \cdots + \mathrm{ST}_n)) \qquad (3.47)$$

$$\leqslant \mathsf{E}\,f(Z) \quad \forall f \in \mathcal{F}^3, \qquad (3.48)$$

$$\mathsf{P}(\hat{V}_{\mathsf{symm},1,p} \geqslant x) \leqslant c_{3,0}\,\mathsf{P}^{\mathsf{LC}}(n^{-1/2}(\mathrm{ST}_1 + \cdots + \mathrm{ST}_n) \geqslant x) \qquad (3.49)$$

$$\leqslant c_{3,0}\,\mathsf{P}(Z \geqslant x) \quad \forall x \in \mathbb{R}; \qquad (3.50)$$

here again the $\mathrm{ST}_i$'s are i.i.d. $\mathsf{ST}(p)$ r.v.'s.

Note that $\operatorname{ess\,sup} \hat{V}_{\mathsf{symm},1,p} = \sqrt{n/p}$, while $\operatorname{ess\,sup} V = \sqrt{n}$. Thus, the tails of $\hat{V}_{\mathsf{symm},1,p}$ are longer in some sense than those of $V$, so that the symmetry test based on an inequality such as (3.49) or (3.50) may be less conservative and hence more powerful than the corresponding test based on an inequality such as (3.25), especially if the size of the test is small enough. An interesting question is how the generalized moments $\mathsf{E}\,f(\hat{V}_{\mathsf{symm},m,p})$ compare with $\mathsf{E}\,f(V)$ for $f \in \mathcal{F}^3$.

It may also be of interest to compare $\mathsf{E}\,f(\hat{V}_{\mathsf{symm},m,p})$ with $\mathsf{E}\,f(V_{\mathsf{symm},m,p})$ (assuming that $p_i = p$ for all $i$). It seems that neither of them dominates the other one in general. In view of (3.43), it seems likely that $\mathsf{E}\,f(\hat{V}_{\mathsf{symm},m,p})$ will be greater than $\mathsf{E}\,f(V_{\mathsf{symm},m,p})$ if the tails of the distributions of the $X_i$'s are not too heavy: then the variability of the $|\hat{X}_i|$'s will be less than that of the

$|X_i|$'s and hence, heuristically, the tails of $\hat{V}_{\mathsf{symm},m,p}$ will be heavier than those of $V_{\mathsf{symm},m,p}$.

Anyway, a definite advantage of $\hat{V}_{\mathsf{symm},m,p}$ (over $V$ and $V_{\mathsf{symm},m,p}$) is that the denominator of its ratio expression in (3.44) is nonzero a.s. On the other hand, an obvious disadvantage of $\hat{V}_{\mathsf{symm},m,p}$ is that to compute its value one needs to know the distributions of the $X_i$'s. In statistical practice, $\hat{V}_{\mathsf{symm},m,p}$ may be approximated (at least in the case when the $X_i$'s are i.i.d.) by replacing $\tilde{X}_i$ for each $i$ by $X_{J(i)}$, where $J(1), \ldots, J(n)$ are (conditionally, given $X_1, \ldots, X_n$) i.i.d. r.v.'s, each $J(i)$ having the uniform distribution on the set

$$\{j \in \{1, \ldots, n\} \colon X_j \neq 0\};$$

this set is nonempty with a probability close to 1 if $p$ is close to 1 or $n$ is large; however, if this set happens to be empty, one can just set $\tilde{X}_i := 0$, so that one has $\hat{X}_i = 0$. It would be interesting to compare generalized moments and tails of this "practical" version of $\hat{V}_{\mathsf{symm},m,p}$ with those of $\hat{V}_{\mathsf{symm},m,p}$ itself.

## 4. Proofs

### *4.1. Statements of lemmas and proofs of the main results*

The proofs of the main results are preceded in this subsection by some definitions and a series of lemmas. At least one of them (Lemma 4.1.12) may be of independent interest. The proofs of the lemmas are deferred further to Subsection 4.2.

Let us introduce more classes of functions, in addition to the classes $\mathcal{H}_+^3$, $\mathcal{F}_+^3$, $\mathcal{F}_-^3$, $\mathcal{F}^3$, and $\mathcal{G}^3$ (recall (1.3), (2.1), (2.13), (2.14), and (2.15):

$$\mathcal{G}_+^3 := \{f \colon \exists a \in \mathbb{R}, b \in \mathbb{R}, h \in \mathcal{H}_+^3 \ \forall x \in \mathbb{R} \ f(x) = a + b\,x + h(x)\}; \qquad (4.1)$$

$$\mathcal{G}_{++}^3 := \{f \colon \exists a \in \mathbb{R}, b \geqslant 0, h \in \mathcal{H}_+^3 \ \forall x \in \mathbb{R} \ f(x) = a + b\,x + h(x)\}. \qquad (4.2)$$

*Remark* 4.1. It is not difficult to see that, if a function $f$ is in $\mathcal{F}_+^3$ or any other defined above class of functions, then the shifted function $x \mapsto f(x + a)$ is also in the same class, for any real constant $a$. That is, all these classes of functions are shift-invariant.

**Lemma 4.1.1.** *Suppose that a function $g \colon \mathbb{R} \to \mathbb{R}$ is convex and such that there exists a finite limit $g(-\infty) := \lim_{x \to -\infty} g(x)$; in particular, the latter condition will obviously be the case if $g$ is nonnegative and nondecreasing. Then $g'(-\infty) = 0$, where $g'$ is the right derivative of $g$.*

**Lemma 4.1.2.** *If $\alpha$ is a natural number then $\mathcal{H}_+^\alpha$ coincides with the class $\tilde{\mathcal{H}}_+^\alpha$ of all functions $f \colon \mathbb{R} \to \mathbb{R}$ such that the derivative $f^{(\alpha-1)}$ is everywhere finite and convex, and $f^{(0)}(-\infty) = \cdots = f^{(\alpha-1)}(-\infty) = 0$. Moreover, if $f \in \mathcal{H}_+^\alpha$, then all the functions $f^{(0)}, \ldots, f^{(\alpha-1)}$ are nonnegative.*

**Lemma 4.1.3.** *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function such that $f''$ is finite, nonnegative, nondecreasing, and convex, with $f''(-\infty) = 0$. Then $f \in \operatorname{cl}\mathcal{G}_+^3$. If, moreover, $f$ is nondecreasing, then $f \in \operatorname{cl}\mathcal{G}_{++}^3$.*

**Lemma 4.1.4.** *One has $\mathcal{F}_+^3 = \operatorname{cl}\mathcal{G}_{++}^3$.*

**Lemma 4.1.5.** *One has $\mathcal{F}^3 = \operatorname{cl}\mathcal{G}^3$, where $\mathcal{G}^3$ is defined by (2.15).*

**Lemma 4.1.6.** *If $f \in \mathcal{F}_+^3$, then either $f(x) = O(x)$ as $x \to \infty$ or $\liminf_{x\to\infty} f(x)/x^2 \in (0, \infty]$.*

**Lemma 4.1.7.** *If $f \in \mathcal{F}_+^3$, then $f(x) = O(|x|)$ as $x \to -\infty$.*

**Lemma 4.1.8.** *$\mathcal{F}^3 \neq \mathcal{G}^3$.*

**Proposition 4.2.** *There exists a function $g \in \mathcal{F}_+^3 \setminus \mathcal{G}_+^3$. Since $\mathcal{G}_{++}^3 \subseteq \mathcal{G}_+^3$, it follows that $\mathcal{F}_+^3 \neq \mathcal{G}_{++}^3$.* (This proposition complements Lemmas 4.1.4 and 4.1.8; it will not be used elsewhere in this paper.)

The following two lemmas are essentially well known. Their statements (and proofs) are given here for easy reference.

**Lemma 4.1.9.** (Cf. e.g. [20] and [4, Lemma 4.3].) *Let $X$ be a r.v. such that $\mathsf{E}\,X \leqslant 0$ and $-a \leqslant X \leqslant b$ a.s. for some positive real numbers $a$ and $b$. Let $\mathrm{BS} \sim \mathrm{BS}(p)$ with*

$$p := \frac{a}{b+a}.$$

*Then*

$$\mathsf{E}\,f(X) \leqslant \mathsf{E}\,f(\sqrt{a\,b}\,\mathrm{BS})$$

*for any nondecreasing convex function $f$, and hence for any function $f \in \mathcal{F}_+^3$.*

**Lemma 4.1.10.** *If $X$ is a zero-mean r.v., then $\mathsf{E}\,f(c\,X)$ is nondecreasing in $c \geqslant 0$ for any convex function $f$ and hence for any $f \in \mathcal{F}_+^3$.*

**Lemma 4.1.11.** *Let $\mathrm{BS}(p) \sim \mathrm{BS}(p)$. Then $e(p) := \mathsf{E}\,f(\mathrm{BS}(p))$ is nonincreasing in $p \in (0,1)$ for any $f \in \mathcal{H}_+^2$ and so, by (1.4), for any $f \in \mathcal{H}_+^3$, whence, by Lemma 4.1.4, for any $f \in \mathcal{F}_+^3$.*

The extension from Theorem 2.2 to Theorem 2.4 to Theorem 2.5 is based in part on the following simple lemma, which may be of independent interest.

**Lemma 4.1.12.** *Suppose that for every $i \in \{1, \ldots, n\}$ one has (2.8) and (2.9). Then*

$$\mathsf{E}\,f(S_n) \leqslant \mathsf{E}\,f(c_1\mathrm{BS}_1 + \cdots + c_n\mathrm{BS}_n) \tag{4.3}$$

*for any $f \in \mathcal{H}_+^2$, and so, by (1.4), for any $f \in \mathcal{H}_+^3$, whence, by Lemma 4.1.4 and Lebesgue's dominated convergence theorem, for any $f \in \mathcal{F}_+^3$.*

For $m \geqslant 1$, introduce

$$
\begin{aligned}
p_* := p_*(m) &:= \frac{2m + 1 - \sqrt{4(m-1)(m+2)+1}}{4(2m-1)} \\
&= \frac{2}{(2m-1)\left(2m+1+\sqrt{4(m-1)(m+2)+1}\right)},
\end{aligned}
\tag{4.4}
$$

so that $p_* \in (0, \frac{1}{2}]$. Introduce also

$$
\begin{aligned}
\delta_1(u,c,p,m) &:= 2c(1 - c^{2m-2})u + 2pc(1 - c^{2m-1}) + c^2(1 - c^{2m-3}); \\
\delta_2(u,c,p,m) &:= (1-p)(1 - c^{2m-1})u^2 \\
&\quad + 2c(1 - c^{2m-2})u + 2pc(1 - c^{2m-1}) + c^2(1 - c^{2m-3}); \\
\delta_3(u,c,p,m) &:= -c^{2m-1}u^2 - 2(c^{2m-1} - cp + c^{2m}p)\,u \\
&\quad + \left((2c + c^2 - 2c^{2m} - c^{2m+1})p - c^{2m-1}\right); \\
\delta_4(u,c,p,m) &:= (1 - c^{2m-1})\,p\,(1 + c + u)^2.
\end{aligned}
$$

**Lemma 4.1.13.** *For any given pair $(p, m)$ such that $m \geqslant 1$ and $p \in (0, 1)$, the following two statements are equivalent to each other.*

(i) *(Cf. Statement (V) in Theorem 2.6.) For every every function $f \in \mathcal{H}_+^3$, the function*

$$
[0, \infty)^n \ni (a_1, a_2) \longmapsto \mathsf{E}\, f(a_1^{1/(2m)}\mathrm{BS}_1 + a_2^{1/(2m)}\mathrm{BS}_2),
$$

*is Schur-concave, where $\mathrm{BS}_i \overset{\text{i.i.d.}}{\sim} \mathsf{BS}(p)$, $i = 1, 2$.*

(ii) *for all $u \in \mathbb{R}$ and $c \in (0, 1)$, one has the inequalities*

$$
\begin{aligned}
\delta_1(u,c,p,m)\,\boldsymbol{I}\{u \geqslant 0\} &\geqslant 0; \\
\delta_2(u,c,p,m)\,\boldsymbol{I}\{-c \leqslant u \leqslant 0\} &\geqslant 0; \\
\delta_3(u,c,p,m)\,\boldsymbol{I}\{-1 \leqslant u \leqslant -c\} &\geqslant 0.
\end{aligned}
$$

**Lemma 4.1.14.** *For all $u \geqslant 0$, $c \in (0, 1)$, $p \in [p_*, 1)$, and $m \geqslant 1$, one has $\delta_1(u) := \delta_1(u, c, p, m) \geqslant 0$.*

**Lemma 4.1.15.** *For all $u \in [-c, 0]$, $c \in (0, 1)$, $p \in [p_*, 1)$, and $m \geqslant 1$, one has $\delta_2(u) := \delta_2(u, c, p, m) \geqslant 0$.*

**Lemma 4.1.16.** *For all $u \in [-1, -c]$, $c \in (0, 1)$, $p \in [p_*, 1)$, and $m \geqslant 1$, one has $\delta_3(u) := \delta_3(u, c, p, m) \geqslant 0$.*

**Lemma 4.1.17.** *(Recall (2.3) and (4.4).) For $p \in (0, 1)$ and $m \geqslant 1$, one has*

$$
m \geqslant m_*(p) \iff p \geqslant p_*(m).
$$

**Lemma 4.1.18.** *In the context of Theorem 2.6, implication (III) $\implies$ (I) is true.*

*Proof of Theorem 2.6.* It suffices to prove the implications

$$
\begin{aligned}
\text{(I)} &\implies \text{(V)} \implies \text{(IV)} \implies \text{(II)} \implies \text{(III)} \implies \text{(I)} \quad \text{and} \\
&\text{(II)} \implies \text{(VII)} \implies \text{(VI)} \implies \text{(II)}.
\end{aligned}
\tag{4.5}
$$

(I) $\implies$ (V): Suppose that condition $m \geqslant m_*(p)$ of item (I) takes place. By Lemma 4.1.17, this condition is equivalent to $p \geqslant p_*(m)$. Now statement (V) with $\mathcal{H}_+^3$ in place of $\mathcal{F}_+^3$ follows from Lemmas 4.1.13, 4.1.14, 4.1.15, and 4.1.16. Hence, by the definition (4.2) of $\mathcal{G}_{++}^3$, one has (V) with $\mathcal{G}_{++}^3$ in place of $\mathcal{F}_+^3$. To complete the proof of implication (I) $\implies$ (V), it remains to use Lemma 4.1.4; recall that the r.v.'s $\mathsf{BS}_i$ each take on only finitely many (namely, two) values.

(V) $\implies$ (IV): By the well-known result by Muirhead [24] (see, e.g., [23, Remark B.1 of Chapter 2]), a function of $n$ nonnegative arguments is Schur-concave iff it is Schur-concave in any two of its arguments. Now implication (V) $\implies$ (IV) follows in view of Remark 4.1 on page 25, by conditioning on all of the r.v.'s $\mathsf{BS}_1, \ldots, \mathsf{BS}_n$ except any given two of them.

(IV) $\implies$ (II): Let here $a_1 := c_1^{2m}, \ldots, a_n := c_n^{2m}$ and $\overline{a} := (a_1 + \cdots + a_n)/n$, so that $s^{(m)} = \overline{a}^{\,1/(2m)}$. Note that $(a_1, \ldots, a_n) \succcurlyeq (\underbrace{\overline{a}, \ldots, \overline{a}}_{n})$. Now implication (IV) $\implies$ (II) follows.

(II) $\implies$ (III): This implication is trivial.

(III) $\implies$ (I): This implication is true by Lemma 4.1.18.

(II) $\implies$ (VII): This implication follows immediately from Lemma 4.1.12.

(VII) $\implies$ (VI): This implication is trivial.

(VI) $\implies$ (II): This implication is also trivial. $\qquad\square$

*Proof of Theorem 2.7.* Replace all functions $f$ in Theorem 2.6 with their reflections $\tilde{f}$ defined by $\tilde{f}(x) := f(-x)$ for all real $x$, replace all r.v.'s $X$ there with $-X$, and interchange $p$ with $q$, $a_i$ with $b_i$, and $A_{i-1}$ with $B_{i-1}$. Note that (i) $\tilde{f}(-X) = f(X)$; (ii) $\mathsf{E}\,X \leqslant 0 \iff \mathsf{E}(-X) \geqslant 0$; (iii) $X \sim \mathsf{BS}(p) \iff -X \sim \mathsf{BS}(q)$; and (iv) $(S_0, \ldots, S_n)$ is a supermartingale with $S_0 \leqslant 0$ a.s. iff $(-S_0, \ldots, -S_n)$ is a submartingale with $-S_0 \geqslant 0$ a.s. Thus, Theorem 2.7 follows immediately from Theorem 2.6. $\qquad\square$

*Proof of Proposition 2.9.* This follows immediately from Lemmas 4.1.5 and 4.1.8. $\qquad\square$

*Proof of Theorem 2.10.* It suffices to prove the same implications, (4.5), as in the proof of Theorem 2.6, only with the changes stated in the formulation of Theorem 2.10. Below, all these implications are understood in the context of Theorem 2.10. The proofs of most of these implications are similar to their proofs in the context of Theorem 2.6. Below, only the most significant changes are described.

(I) $\implies$ (V): To prove this implication, in view of Theorems 2.6 and 2.7 and Proposition 2.9, it suffices to verify that the function (2.12) is Schur-concave when $f(x) = x^2$ (for all real $x$) and $n = 2$. Thus, it suffices to verify that, for

any given $m \geqslant 1$, the expression

$$\mathsf{E}(\mathrm{BS}_1 \cos^{1/m} \theta + \mathrm{BS}_2 \sin^{1/m} \theta)^2 = \cos^{2/m} \theta + \sin^{2/m} \theta$$

is nondecreasing in $\theta \in [0, \pi/4]$. But this is easy to see.

(III) $\implies$ (I): This implication follows from Theorems 2.6 and 2.7 and the observation that both classes $\mathcal{F}_+^3$ and $\mathcal{F}_-^3$ are contained in $\mathcal{F}^3$.

(II) $\implies$ (VII): In view of Theorems 2.6 and 2.7, Proposition 2.9, and Lebesgue's dominated convergence theorem, it suffices to verify that inequality (2.11) holds when $f(x) = x^2$ for all real $x$ and $(S_0, \ldots, S_n)$ is a martingale as described in the formulation of Theorem 2.10. Note that inequality (4.3) holds for the function $f_0(x) := x_+^2$ in place of $f$, since $f_0 \in \mathcal{G}_+^2$. It also holds for the function $\tilde{f}_0(x) := (-x)_+^2$ in place of $f$, given that $(S_0, \ldots, S_n)$ is a martingale as described. Thus, (4.3) holds when $f(x) = x^2$ for all real $x$. It remains to note that

$$\mathsf{E}(c_1 \mathrm{BS}_1 + \cdots + c_n \mathrm{BS}_n)^2 = n\,(s^{(1)})^2 \leqslant n\,(s^{(m)})^2 = \mathsf{E}(s^{(m)} \cdot (\mathrm{BS}_1 + \cdots + \mathrm{BS}_n))^2,$$

so that one does have inequality (2.11) when $f(x) = x^2$ for all real $x$. $\qquad\square$

### 4.2. Proofs of the lemmas

*Proof of Lemma 4.1.1.* The convexity of $g$ implies $g(0) - g(x) \geqslant g'(x)(-x)$ and $g(2x) - g(x) \geqslant g'(x)x$, so that $|g'(x)||x| \leqslant \max(|g(0) - g(x)|, |g(2x) - g(x)|)$ for all $x \in \mathbb{R}$. Letting now $x \to -\infty$ and using the existence of the finite limit $g(-\infty)$, one has $g'(-\infty) = 0$. $\qquad\square$

*Proof of Lemma 4.1.2.* This lemma was stated essentially as Proposition 1.1 in [31]. The proof given here is a little more detailed. Assume first that $f \in \mathcal{H}_+^\alpha$, so that $f(x) = \int (x - t)_+^\alpha \, d\mu(t)$ for a function $\mu \in \mathcal{M}_+^\alpha$, whence $f^{(\alpha-1)}(x) = \alpha! \int (x - t)_+ \, d\mu(t)$ is convex as a limit of linear combinations with nonnegative coefficients of convex functions $x \mapsto (x - t)_+$. The conditions $f^{(0)}(-\infty) = \cdots = f^{(\alpha-1)}(-\infty) = 0$ follow by Lebesgue's dominated convergence theorem. Thus, $f \in \tilde{\mathcal{H}}_+^\alpha$. Moreover, it is clear that all the functions $f^{(0)}, \ldots, f^{(\alpha-1)}$ are nonnegative.

Assume now that $f \in \tilde{\mathcal{H}}_+^\alpha$. Consider first the case $\alpha = 1$. Then $f$ is convex and $f(-\infty) = 0$. Hence, by Lemma 4.1.1, one has $f'(-\infty) = 0$. Therefore,

$$f(x) = \int_{-\infty}^x f'(u)\, du = \int_{-\infty}^x du \int_{-\infty}^u d\,f'(v) = \int (x - v)_+ \, d\,\mu(v),$$

by Fubini's theorem, where $\mu := f'$; thus, $f \in \mathcal{H}_+^1$. The case of any natural $\alpha \geqslant 2$ can now be treated by induction, in a similar manner. Indeed, if $f \in \tilde{\mathcal{H}}_+^\alpha$ for a natural $\alpha \geqslant 2$, then $f' \in \tilde{\mathcal{H}}_+^{\alpha-1}$, by the definition of $\tilde{\mathcal{H}}_+^\alpha$. Hence, for a function $\mu \in \mathcal{M}_+^\alpha$,

$$f(x) = \int_{-\infty}^x f'(u)\, du = \int_{-\infty}^x du \int (u - v)_+^{\alpha-1} \, d\mu(v) = \int (x - v)_+^\alpha \, d\,\mu(v)/\alpha,$$

so that $f \in \mathcal{H}_+^\alpha$. $\qquad\square$

*Proof of Lemma 4.1.3.* For the given function $f$ and any $y \in \mathbb{R}$, introduce the functions defined by the formulas

$$f_{2,y}(x) := \left(f''(y) + f'''(y)(x - y)\right)_+ \boldsymbol{I}\{x \leqslant y\} + f''(x)\,\boldsymbol{I}\{x > y\}; \qquad (4.6)$$

$$f_y(x) := \left(f(y) + f'(y)(x - y) + \int_{-\infty}^y f_{2,y}(u)\,(u - x)_+\,\mathrm{d}u\right)\boldsymbol{I}\{x \leqslant y\} \qquad (4.7)$$
$$+ f(x)\,\boldsymbol{I}\{x > y\}$$

for all real $x$. Here $f'''$ denotes the right derivative of the convex function $f''$, so that $f'''$ is nondecreasing. Note that

$$f_y'' = f_{2,y}. \qquad (4.8)$$

Since $f''$ is convex, one has

$$f''(x) \geqslant f''(y) + f'''(y)(x - y) \qquad (4.9)$$

for all $y$ and $x$; also, it is given that $f''$ is nonnegative; it follows that

$$f'' \geqslant f_{2,y}. \qquad (4.10)$$

Observe that, moreover, the family of functions $(f_{2,y})$ is nonincreasing in $y \in \mathbb{R}$. Indeed, let $y$ and $y_1$ be any real numbers such that $y_1 < y$. Then $f_{2,y_1} = f'' \geqslant f_{2,y}$ on $[y_1, \infty)$, in view of (4.6) and (4.10). Recalling (4.9) and the fact that $f'''$ is nondecreasing, one has the inequalities $f''(y_1) \geqslant f''(y) + f'''(y)(y_1 - y)$ and $f'''(y_1)(x - y_1) \geqslant f'''(y)(x - y_1)$ for all $x \leqslant y_1$; adding these inequalities, one sees that $f''(y_1) + f'''(y_1)(x - y_1) \geqslant f''(y) + f'''(y)(x - y)$. It follows, in view of (4.6), that $f_{2,y_1} \geqslant f_{2,y}$ on the interval $(-\infty, y_1]$ as well, and hence on the entire real line.

Using integration-by-parts/Fubini's theorem as in the proof of Lemma 4.1.2, one can verify that for any function $g \in \mathcal{C}^2$ and all real $y$ and $x$

$$g(x) = \left(g(y) + g'(y)(x - y) + \int_{-\infty}^y g''(u)\,(u - x)_+\,\mathrm{d}u\right)\boldsymbol{I}\{x \leqslant y\} \qquad (4.11)$$
$$+ g(x)\,\boldsymbol{I}\{x > y\}.$$

By (4.8), for any real $w$ one has $f_w \in \mathcal{C}^2$, so that one can substitute $f_w$ for $g$ in (4.11). In fact, let us do so for $w \in \{y, y_1\}$, again assuming that $y_1 < y$. At that, by (4.7), one has $f_{y_1} = f = f_y$ on the interval $[y, \infty)$ and hence $f_{y_1}(y) = f_y(y)$ and $f_{y_1}'(y) = f_y'(y)$. Now, since the family of functions $(f_y'')_{y \in \mathbb{R}} = (f_{2,y})_{y \in \mathbb{R}}$ is nonincreasing, one can see that the family $(f_y)_{y \in \mathbb{R}}$ is nonincreasing as well. Next, since $\boldsymbol{I}\{x \leqslant y\} \to 0$ and $\boldsymbol{I}\{x > y\} \to 1$ for each $x$ as $y \to -\infty$, one concludes, in view of (4.8), that for any decreasing sequence $(y_n)$ in $\mathbb{R}$ converging to $-\infty$ one has $f_{y_n} \to f$, in the sense of Definition 2.8.

It remains to verify that for every real $y$ one has $f_y \in \mathcal{G}_+^3$ and, moreover, $f_y \in \mathcal{G}_{++}^3$ in the case when $f$ is known to be nondecreasing. Observe that

$$f_{2,y}(x) = f'''(y)\,(x - z)\,\boldsymbol{I}\{z \leqslant x \leqslant y\} + f''(x)\,\boldsymbol{I}\{x > y\}, \qquad (4.12)$$

where
$$z := y \, \boldsymbol{I}\{f'''(y) = 0\} + (y - f''(y)/f'''(y)) \, \boldsymbol{I}\{f'''(y) \neq 0\}.$$

Indeed, $f'''$ is nonnegative and nondecreasing (since $f''$ is nondecreasing and convex). Hence, in the case when $f'''(y) = 0$, one has $f''' = 0$ on the entire interval $(-\infty, y]$. This and the condition $f''(-\infty) = 0$ implies $f'' = 0$ on the entire interval $(-\infty, y]$, so that $f''(y) = 0$. Now one sees that expressions (4.6) and (4.12) both equal $f''(x) \, \boldsymbol{I}\{x > y\}$ in the case when $f'''(y) = 0$. In the other case, when $f'''(y) \neq 0$, one has $f'''(y) > 0$ (since $f'''$ is nonnegative). Also, here $f''(y) + f'''(y)(x - y) = f'''(y) \, (x - z)$, whence (4.12) again follows.

Now, for the right derivative $f'_{2,y}$ of $f_{2,y}$, (4.12) yields
$$f'_{2,y}(x) = f'''(y) \, \boldsymbol{I}\{z \leqslant x \leqslant y\} + f'''(x) \, \boldsymbol{I}\{x > y\}.$$

Since $f'''$ is nonnegative and nondecreasing, it follows now that $f'_{2,y}$ is nondecreasing. Therefore, $f_{2,y}$ is convex. That is, by (4.8), $f''_y$ is convex. Also, (4.12) and (4.8) show that $f''_y = 0$ on the interval $(-\infty, z]$. This means that

$$f_y(x) = a + b \, x \text{ for some real constants } a \text{ and } b \text{ and all } x \leqslant z. \qquad (4.13)$$

Let now
$$h_y(x) := f_y(x) - (a + b \, x)$$

for all real $x$. Then $h''_y = f''_y$ is convex. Moreover, $h_y = 0$ on the interval $(-\infty, z]$, so that $h_y(-\infty) = h'_y(-\infty) = h''_y(-\infty) = 0$. By Lemma 4.1.2, $h_y \in \mathcal{H}^3_+$. Thus, $f_y \in \mathcal{G}^3_+$.

If, moreover, $f$ is nondecreasing, then $f' \geqslant 0$. Hence and because (in view of (4.10)) $f''(u) \geqslant f_{2,y}(u) \geqslant f_{2,y}(u) \, \boldsymbol{I}\{u > x\}$ for all $x$, $y$, and $u$, one has

$$f'_y(x) = \left( f'(y) - \int_{-\infty}^y f_{2,y}(u) \, \boldsymbol{I}\{u > x\} \, \mathrm{d}u \right) \boldsymbol{I}\{x \leqslant y\} + f'(x) \, \boldsymbol{I}\{x > y\}$$
$$\geqslant f'(-\infty) \, \boldsymbol{I}\{x \leqslant y\} + f'(x) \, \boldsymbol{I}\{x > y\} \geqslant 0$$

for all $x$. Now (4.13) implies $b \geqslant 0$. Since $h_y \in \mathcal{H}^3_+$, one finally sees that $f_y \in \mathcal{G}^3_{++}$. $\qquad \square$

*Proof of Lemma 4.1.4.* First note that $\mathrm{cl}\,\mathcal{F}^3_+ = \mathcal{F}^3_+$, because the pointwise convergence preserves both the monotonicity and the convexity.

Next, take any $f \in \mathcal{G}^3_{++}$, so that

$$f(x) = a + b \, x + \int (x - t)^3_+ \, \mathrm{d}\mu(t)$$

for all $x$, where $a \in \mathbb{R}$, $b \geqslant 0$, and $\mu$ is nondecreasing and $\int (-t)^3_+ \, \mathrm{d}\mu(t) < \infty$. It follows that $f$ is nondecreasing and convex, since the functions $x \mapsto (x - t)^3_+$ are so. Similarly, $f''$ is nondecreasing and convex, since $f''(x) = 6 \int (x - t)_+ \, \mathrm{d}\mu(t)$. That is, $f \in \mathcal{F}^3_+$ for any $f \in \mathcal{G}^3_{++}$, so that $\mathcal{G}^3_{++} \subseteq \mathcal{F}^3_+$, whence $\mathrm{cl}\,\mathcal{G}^3_{++} \subseteq \mathrm{cl}\,\mathcal{F}^3_+ = \mathcal{F}^3_+$.

It remains to show that $\mathcal{F}^3_+ \subseteq \mathrm{cl}\,\mathcal{G}^3_{++}$. Take any $f \in \mathcal{F}^3_+$. Then, by definition (2.1), $f$ and $f''$ are nondecreasing and convex. Hence, $f'$ is nonnegative, nondecreasing, and convex. Now Lemma 4.1.1 yields $f''(-\infty) = 0$. Also, $f''$ is nonnegative, since $f$ is convex. Thus, by Lemma 4.1.3, $f \in \mathrm{cl}\,\mathcal{G}^3_{++}$. $\qquad \square$

*Proof of Lemma 4.1.5.* First note that $\operatorname{cl}\mathcal{F}^3 = \mathcal{F}^3$, because the pointwise convergence preserves the convexity.

Next, it is trivial that $\mathcal{G}^3 \subseteq \mathcal{F}^3$, whence $\operatorname{cl}\mathcal{G}^3 \subseteq \operatorname{cl}\mathcal{F}^3 = \mathcal{F}^3$.

It remains to show that $\mathcal{F}^3 \subseteq \operatorname{cl}\mathcal{G}^3$. Take any $f \in \mathcal{F}^3$. Then, by definition (2.14), $f$ and $f''$ are convex. The latter condition implies that at least one of the following three cases must take place: $f''$ is nondecreasing on $\mathbb{R}$ or $f''$ is nonincreasing on $\mathbb{R}$ or $f''$ switches from nonincreasing to nondecreasing.

*Case 1: $f''$ is nondecreasing on $\mathbb{R}$.* Since $f$ is convex, $f'' \geqslant 0$ on $\mathbb{R}$. Hence, there exists the limit $c := f''(-\infty) \in [0, \infty)$. Let

$$g(x) := f(x) - c\,x^2/2$$

for all real $x$. Then $g'' = f'' - c = f'' - f''(-\infty) \geqslant 0$, since $f''$ is nondecreasing. Also, $g'' = f'' - c$ is nondecreasing and convex, since $f''$ is so. In addition, $g''(-\infty) = 0$. Therefore, by Lemma 4.1.3, $g \in \operatorname{cl}\mathcal{G}_+^3$. That is, there exists a sequence of functions $(g_n)$ such that $g_n \to g$ and

$$g_n(x) = a_n + b_n\,x + h_n(x)$$

for all real $x$, where, for each $n$, $a_n$ and $b_n$ are real constants and $h_n$ is a function in $\mathcal{H}_+^3$; then $h_n \in \mathcal{F}_+^3$ (because $\mathcal{H}_+^3 \subseteq \mathcal{F}_+^3$, as seen, for example, from the proof of Lemma 4.1.2). Let now

$$f_n(x) := c\,x^2/2 + g_n(x) = c\,x^2/2 + a_n + b_n\,x + h_n(x)$$

for all $x$ and $n$. Then $g_n \to g$ implies $f_n \to f$. Moreover, for every $n$ one has $f_n \in \mathcal{G}^3$. Indeed, if $b_n < 0$, then the function $h_n$ belongs to $\mathcal{F}_+^3$ and the function $x \mapsto a_n + b_n\,x$ belongs to $\mathcal{F}_-^3$; and if $b_n \geqslant 0$, then the function $x \mapsto a_n + b_n\,x + h_n(x)$ belongs to $\mathcal{F}_+^3$ and the function $x \mapsto 0$ belongs to $\mathcal{F}_-^3$. Thus, $f \in \operatorname{cl}\mathcal{G}^3$ for any $f \in \mathcal{F}^3$ satisfying the condition of Case 1.

*Case 2: $f''$ is nonincreasing on $\mathbb{R}$.* This case reduces to Case 1 by considering the function $x \mapsto \tilde{f}(x) := f(-x)$ in place of $f$. Indeed, if $f \in \mathcal{F}^3$ and $f''$ is nonincreasing, then $\tilde{f} \in \mathcal{F}^3$ and $\tilde{f}''$ is nondecreasing. Moreover, $f_n \in \mathcal{G}^3 \iff \tilde{f}_n \in \mathcal{G}^3$, where $\tilde{f}_n(x) := f_n(-x)$ for all real $x$.

*Case 3: There exists some real $x_0$ such that $f''$ is nonincreasing on $(-\infty, x_0]$ and nondecreasing on $[x_0, \infty)$.* Here without loss of generality (w.l.o.g.) $x_0 = 0$. Let

$$g(x) := f(x) - f'(0)x - f''(0)x^2/2$$

for all real $x$, so that $g(0) = f(0)$ and $g'(0) = g''(0) = 0$; moreover, $g'' = f'' - f''(0)$ is convex on $\mathbb{R}$ (since $f''$ is so), nonincreasing on $(-\infty, 0]$, and nondecreasing on $[0, \infty)$, whence $g'' \geqslant 0$. Let, for all real $x$,

$$h_+(x) := g''(x)\boldsymbol{I}\{x > 0\} \quad \text{and} \quad h_-(x) := g''(x)\boldsymbol{I}\{x \leqslant 0\},$$

so that $h_+ + h_- = g''$, $h_\pm \geqslant 0$, $h_+$ is nondecreasing, and $h_-$ is nonincreasing. Also, $h_+$ and $h_-$ are convex, since $g''$ is convex and $g'' \geqslant g''(0) = 0$. Let further, for all real $x$,

$$H_+(x) := \int (x - t)_+\, h_+(t)\, \mathrm{d}t \quad \text{and} \quad H_-(x) := \int (t - x)_+\, h_-(t)\, \mathrm{d}t,$$

so that $H_{\pm}(0) = 0$,

$H'_+(x) = \int \boldsymbol{I}\{x > t\}\, h_+(t)\, \mathrm{d}t = \int_0^x g''(t)\, \mathrm{d}t\, \boldsymbol{I}\{x > 0\} = g'(x)\boldsymbol{I}\{x > 0\}\,;$
$H''_+(x) = g''(x)\boldsymbol{I}\{x > 0\} = h_+(x);$

$H'_-(x) = -\int \boldsymbol{I}\{t > x\}\, h_-(t)\, \mathrm{d}t = -\int_x^0 g''(t)\, \mathrm{d}t\, \boldsymbol{I}\{x \leqslant 0\} = g'(x)\boldsymbol{I}\{x \leqslant 0\}\,;$
$H''_-(x) = g''(x)\boldsymbol{I}\{x \leqslant 0\} = h_-(x).$

It follows that $H'_+ + H'_- = g'$, whence for all real $x$ one has $H_+(x) + H_-(x) = g(x) - g(0)$, that is,

$$f(x) = f(0) + f'(0)x + f''(0)x^2/2 + H_+(x) + H_-(x). \qquad (4.14)$$

Also, $H''_+ = h_+$ is nonnegative, nondecreasing, and convex, and hence $H_+$ is also convex. Also, since $h_+ \geqslant 0$, the first expression for $H'_+(x)$ above shows that $H'_+ \geqslant 0$. Thus, $H_+$ and $H''_+$ are nondecreasing and convex; that is, $H_+ \in \mathcal{F}_+^3$. Similarly, $H_- \in \mathcal{F}_-^3$.

Note also that $f''(0) \geqslant 0$, since $f$ belongs to $\mathcal{F}^3$ and is hence convex. If $f'(0) < 0$, then the function $H_+$ belongs to $\mathcal{F}_+^3$ and the function $x \mapsto f(0) + f'(0)x + H_-(x)$ belongs to $\mathcal{F}_-^3$; and if $f'(0) \geqslant 0$, then the function $x \mapsto f(0) + f'(0)x + H_+(x)$ belongs to $\mathcal{F}_+^3$ and the function $H_-$ belongs to $\mathcal{F}_-^3$. Thus, (4.14) implies that $f \in \mathcal{G}^3 \subseteq \operatorname{cl}\mathcal{G}^3$ for any $f \in \mathcal{F}^3$ satisfying the condition of Case 3.

One concludes that, in all cases $f \in \mathcal{F}^3$ implies $f \in \operatorname{cl}\mathcal{G}^3$. That is, $\mathcal{F}^3 \subseteq \operatorname{cl}\mathcal{G}^3$ indeed. □

*Proof of Lemma 4.1.6.* Let $f \in \mathcal{F}_+^3$.
  *Case 1: $f'' = 0$ on $\mathbb{R}$.*  Then there exist real $a$ and $b$ such that $f(x) = a + b\,x$ for all real $x$, so that $f(x) = O(x)$ as $x \to \infty$.
  *Case 2: there exists some $t \in \mathbb{R}$ such that $f''(t) \neq 0$.*  By (2.1), $f''$ is nonnegative (because $f$ is convex) and nondecreasing. Hence, $f''(t) > 0$ and $f''(x) \geqslant f''(t)$ for all $x \geqslant t$. It follows that $f(x) \geqslant f(t) + f'(t)\,(x - t) + f''(t)\,(x - t)^2/2$ for all $x \geqslant t$, whence $\liminf_{x \to \infty} f(x)/x^2 \geqslant f''(t)/2 > 0$. □

*Proof of Lemma 4.1.7.* Let $f \in \mathcal{F}_+^3$. Since $f$ is nondecreasing, one has $f(x) \leqslant f(0)$ for all $x \leqslant 0$. On the other hand, $f(x) \geqslant f(0) + f'(0)x$ for all real $x$, since $f$ convex. It follows that $|f(x)| \leqslant |f(0)| + |f'(0)x|$ for all $x \leqslant 0$, so that $f(x) = O(|x|)$ as $x \to -\infty$. □

*Proof of Lemma 4.1.8.* For all real $x$, let

$$f(x) := \tfrac{8}{3}\,(1 - x)^{3/2}\,\boldsymbol{I}\{x \leqslant 0\} + (\tfrac{8}{3} - 4x + x^2 + \tfrac{1}{6}\,x^3 + \tfrac{1}{16}\,x^4)\,\boldsymbol{I}\{x > 0\}. \quad (4.15)$$

Then it is easy to see that $f \in \mathcal{F}^3$.
  Suppose that $f \in \mathcal{G}^3$. Then, by (2.15), there exist $c \geqslant 0$, $f_+ \in \mathcal{F}_+^3$, and $f_- \in \mathcal{F}_-^3$ such that for all real $x$

$$f(x) = c\,x^2/2 + f_+(x) + f_-(x).$$

Let $x \to -\infty$. Then, by Lemma 4.1.7, $f_+(x) = O(|x|)$, while by Lemma 4.1.6, either $f_-(x) = O(|x|)$ or $\liminf_{x \to -\infty} f_-(x)/x^2 > 0$. It follows that either

$f(x) = O(|x|)$ as $x \to -\infty$ or $\liminf_{x \to -\infty} f(x)/x^2 > 0$. However, neither of these two alternatives is compatible with the fact that $f(x) = \frac{8}{3}(1-x)^{3/2}$ for $x \leqslant 0$. Thus, $f \in \mathcal{F}^3 \setminus \mathcal{G}^3$. $\qquad\square$

*Proof of Proposition 4.2.* Let $g := f'$, where $f$ is defined by (4.15). Then it is easy to see that $g \in \mathcal{F}_+^3$.

Suppose that $g \in \mathcal{G}_+^3$. Then, by (4.1), there exist real $a$ and $b$ and $h \in \mathcal{H}_+^3$ such that for all real $x$ one has $g(x) = a + b\,x + h(x)$ and hence $g'(x) = b + h'(x)$. By Lemma 4.1.2, $h(-\infty) = h'(-\infty) = 0$. Hence, $b = g'(-\infty)$. But, by inspection, $g'(-\infty) = 0$. It follows that $b = 0$, and so, $g(-\infty) = a + h(-\infty) = a \in \mathbb{R}$. This contradicts the fact that $g(-\infty) = -\infty$. $\qquad\square$

*Proof of Lemma 4.1.9.* Write

$$X = \frac{X+a}{b+a}\,b + \frac{b-X}{b+a}\,(-a).$$

Let $f$ be any nondecreasing convex function. The convexity (together with the condition $-a \leqslant X \leqslant b$ a.s.) implies that

$$f(X) \leqslant \frac{X+a}{b+a}\,f(b) + \frac{b-X}{b+a}\,f(-a) = \mathsf{E}\,f(\sqrt{a\,b}\,\mathrm{BS}) + \frac{f(b)-f(-a)}{b+a}\,X$$

a.s. Now, since $f$ is nondecreasing and $\mathsf{E}\,X \leqslant 0$, the lemma follows. $\qquad\square$

*Proof of Lemma 4.1.10.* Since $f$ is convex, the function $[0,\infty) \ni c \mapsto f(c\,X)$ is convex as well. Hence, the function $[0,\infty) \ni c \mapsto g(c) := \mathsf{E}\,f(c\,X)$ is convex. Since $\mathsf{E}\,X = 0$, one has $g(c) \geqslant g(0)$ for all real $c$, by Jensen's inequality. Therefore, the right derivative of $g$ is nonnegative at $0$ and hence on $[0,\infty)$. Now the lemma follows. $\qquad\square$

*Proof of Lemma 4.1.11.* In view of the definition of the class $\mathcal{H}_+^2$, it suffices to verify the statement of the lemma for all functions $f$ of the form $f_t(x) := (x-t)_+^2$, for all real $t$, so that $e(p) = \mathsf{E}(\mathrm{BS}(p)-t)_+^2$. Then

$$e'(p) = \left(\sqrt{\tfrac{q}{p}} - t\right)\left(-\sqrt{\tfrac{p}{q}} - t\right)\,\boldsymbol{I}\left\{-\sqrt{\tfrac{p}{q}} < t < \sqrt{\tfrac{q}{p}}\right\} \leqslant 0.$$

$\qquad\square$

*Proof of Lemma 4.1.12.* W.l.o.g., the $\mathrm{BS}_i$'s are independent of the $S_i$'s. For $i = 0, 1, \ldots, n$ and $f \in \mathcal{H}_+^2$, introduce

$$F_i := \mathsf{E}\,f\left(S_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n\right).$$

Recall that, by Remark 4.1, the classes $\mathcal{H}_+^\alpha$ are invariant with respect to the shifts. Hence, by Lemmas 4.1.9, 4.1.10, and 4.1.11,

$$\begin{aligned}
F_i &= \mathsf{E}\,\mathsf{E}_{i-1}\,f(S_{i-1} + X_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n) \\
&\leqslant \mathsf{E}\,\mathsf{E}_{i-1}\,f(S_{i-1} + \sqrt{A_{i-1}B_{i-1}}\,\tilde{\mathrm{BS}}_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n) \\
&\leqslant \mathsf{E}\,\mathsf{E}_{i-1}\,f(S_{i-1} + c_i\,\tilde{\mathrm{BS}}_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n) \\
&\leqslant \mathsf{E}\,\mathsf{E}_{i-1}\,f(S_{i-1} + c_i\mathrm{BS}_i + c_{i+1}\mathrm{BS}_{i+1} + \cdots + c_n\mathrm{BS}_n) \\
&= F_{i-1}
\end{aligned}$$

for $i = 1, \ldots, n$, where $\mathsf{E}_{i-1}$ denotes the conditional expectation given the $\sigma$-algebra $G_{i-1}$ generated by $H_{i-1}$ and $(\mathrm{BS}_{i+1}, \ldots, \mathrm{BS}_n)$, and the conditional distribution of $\mathrm{B\tilde{S}}_i$ given $G_{i-1}$ is $\mathsf{BS}(p_i)$, with

$$p_i := \frac{A_{i-1}}{B_{i-1} + A_{i-1}},$$

so that $p_i \geqslant p$, according to (2.9). Hence,

$$\mathsf{E}\, f(S_n) = F_n \leqslant F_0 \leqslant \mathsf{E}\, f(c_1 \mathrm{BS}_1 + \cdots + c_n \mathrm{BS}_n);$$

the last inequality follows because $S_0 \leqslant 0$ a.s. and all functions $f \in \mathcal{H}_+^2$ are nondecreasing. $\quad\square$

*Proof of Lemma 4.1.13.* Statement (i) is equivalent to the following: for every $t \in \mathbb{R}$, the function

$$[0, \infty)^2 \ni (a_1, a_2) \longmapsto e_{t,p,m}(a_1, a_2) := \mathsf{E}\, f_t(a_1^{1/(2m)} \mathrm{BS}_1 + a_2^{1/(2m)} \mathrm{BS}_2)$$

is Schur-concave, where $\mathrm{BS}_i \overset{\text{i.i.d.}}{\sim} \mathsf{BS}(p)$ and

$$f_t(x) := \tfrac{1}{3}\,(x - t)_+^3.$$

Using the homogeneity property

$$e_{t,p,m}(\lambda a_1, \lambda a_2) = \lambda^{3/(2m)}\, e_{\lambda^{-1/(2m)}\, t, p, m}(a_1, a_2)$$

for every $\lambda > 0$, one may assume w.l.o.g. that $a_1 + a_2 = 1$, so that $a_1 = \cos^2 \theta$ and $a_2 = \sin^2 \theta$, for some $\theta \in [0, \pi/2]$; moreover, in view of the same homogeneity property, one may replace here the i.i.d. standardized Bernoulli r.v.'s $\mathrm{BS}_1$ and $\mathrm{BS}_2$ with i.i.d. *centered* Bernoulli r.v.'s $\mathrm{BC}_1$ and $\mathrm{BC}_2$, such that

$$\mathsf{P}(\mathrm{BC}_i = 1 - p) = p = 1 - \mathsf{P}(\mathrm{BC}_i = -p), \quad i = 1, 2.$$

Therefore, statement (i) is equivalent to

$$\Delta_{p,m}(\theta, t) := \partial_\theta \left( \mathsf{E}\, f_t(\mathrm{BC}_1\, \cos^{1/m} \theta + \mathrm{BC}_2\, \sin^{1/m} \theta) \right) \tag{4.16}$$

being nonnegative for all $\theta \in [0, \pi/4]$ and all $t \in \mathbb{R}$ (where $\partial_\theta := \partial/\partial\theta$), which is in turn equivalent to

$$\Delta_{1,p,m}(\theta, u) := \frac{m}{(1-p)p}\, \cos^{1-1/m} \theta\, \sin^{1-1/m} \theta$$
$$\times \Delta_{p,m}\big(\theta, -u - p\, \cos^{1/m} \theta - p\, \sin^{1/m} \theta\big)$$
$$= \Delta_{2,p,m}(\cos^{1/m} \theta, \sin^{1/m} \theta, u) \tag{4.17}$$

being nonnegative for all $\theta \in [0, \pi/4]$ and $u \in \mathbb{R}$, where

$$\Delta_{2,p,m}(c_1, c_2, u) :=$$
$$- (c_1^{2m-1} - c_2^{2m-1})\, q\, u_+^2 - (c_2^{2m-1} q + c_1^{2m-1} p)\, (c_1 + u)_+^2$$
$$+ (c_1^{2m-1} q + c_2^{2m-1} p)\, (c_2 + u)_+^2 + (c_1^{2m-1} - c_2^{2m-1})\, p\, (c_1 + c_2 + u)_+^2.$$

Now, in view of the homogeneity relation

$$\Delta_{2,p,m}(c_1, c_2, u) = c_1^{2m+1}\, \Delta_{2,p,m}(1, c_2/c_1, u/c_1), \quad \text{where} \quad 0 < c_2 < c_1,$$

statement (i) reduces to $\Delta_{2,p,m}(1, c, u)$ being nonnegative for all $c \in (0, 1)$ and $u \in \mathbb{R}$.

It remains to note that

$$\Delta_{2,p,m}(1, c, u) = \begin{cases} \delta_1(u, c, p, m) & \text{if} \quad u \geqslant 0, \\ \delta_2(u, c, p, m) & \text{if} \quad u \in [-c, 0], \\ \delta_3(u, c, p, m) & \text{if} \quad u \in [-1, -c], \\ \delta_4(u, c, p, m) & \text{if} \quad u \in [-1 - c, -1], \\ 0 & \text{otherwise}, \end{cases} \qquad (4.18)$$

and $\delta_4(u, c, p, m)$ is manifestly nonnegative for all $c \in (0, 1)$, $u \in \mathbb{R}$, $p \in (0, 1)$, and $m > 1$. $\qquad\square$

*Proof of Lemma 4.1.14.* Note that $\delta_1(u) \geqslant \delta_1(0)$ for $u \geqslant 0$, $c \in (0, 1)$, and $m \geqslant 1$. Next, (4.18) shows that $\delta_1(0) = \delta_2(0)$. Now the lemma follows immediately from Lemma 4.1.15. $\qquad\square$

*Proof of Lemma 4.1.15.* W.l.o.g., $m > 1$. Note that

$$\partial_p \delta_2(u, c, p, m) = (2c - u^2)(1 - c^{2m-1}) > 0$$

for $u \in [-c, 0]$, $c \in (0, 1)$, and $m > 1$, so that w.l.o.g.

$$p = p_* = p_*(m).$$

Next, $\delta_2(u)$ is a convex quadratic polynomial, whose minimum over all $u \in \mathbb{R}$ is attained at

$$u = u_*(c, p, m) := -\frac{c(1 - c^{2m-2})}{(1 - c^{2m-1})(1 - p)}.$$

Hence, it suffices to show that $\gamma(c, p_*)$ is nonnegative for all $c \in [0, 1]$ and $m > 1$, where

$$\gamma(c, p) := \delta_2(u_*(c, p, m), c, p, m)\, \frac{(1 - c^{2m-1})(1 - p)}{c}$$
$$= -2\big(1 - c^{2m-1}\big)^2 p^2 + (1 - c^{2m-1})(2 - c + c^{2m-2} - 2c^{2m-1})p - c^{2m-2}(1 - c)^2$$

The main idea in the proof of this lemma is to replace here the entry of $p^2$ with the equivalent (for $p = p_*$), first-degree in $p$ polynomial expression according to the identity

$$p_*^2 = \frac{(4m^2 - 1)p_* - 1}{2(2m - 1)^2}, \qquad (4.19)$$

to obtain

$$(2m - 1)^2 \gamma(c, p_*) = f(c, p_*),$$

where

$$f(c, p) := (1 - c^{2m-1})^2 - (2m-1)^2(1-c)^2 c^{2m-2}$$
$$+ p(1 - c^{2m-1})(2m-1)\big((2m-3)(1 - c^{2m-1}) - (2m-1)c(1 - c^{2m-3})\big).$$

It suffices to show that $f(c, p) \geqslant 0$ for all $p \in (0, 1)$, $m > 1$, and $c \in (0, 1)$. Introduce

$$g(c) := f(c, p)/c^{2m};$$
$$g_1(c) := g'(c)/c^{2m-3};$$
$$g_2(c) := g_1'(c)\, c^2;$$
$$g_3(c) := g_2'(c)/c^{1-2m};$$
$$g_4(c) := g_3'(c)/c^{1-2m};$$
$$g_5(c) := g_4'(c)/c^{2m-4}.$$

Then, letting

$$s := m - 1 > 0,$$

one has $g_5'(c) = 8c^{-1-2s}\, s\, (1 + s)(1 + 2s)^2(1 + 4s)\big(1 - p + 4s^2 p\big) > 0$ for all $c \in (0, 1)$, and so, $g_5$ is increasing on $(0, 1)$ to

$$g_5(1) = -16\, s\, (1 + s)(1 + 2s)^2\big(1 - p + s + 4s^2 p\big) < 0.$$

Hence, $g_5 < 0$ on $(0, 1)$, so that $g_4$ is decreasing on $(0, 1)$ to

$$g_4(1) = 8\, s\, (1 + 2s)^2\big(1 - p + s + 4s^2 p\big) > 0.$$

Hence, $g_4 > 0$ on $(0, 1)$. Since $g_3(1) = g_2(1) = g_1(1) = g(1) = 0$, it follows successively that $g_3 < 0$, $g_2 > 0$, $g_1 < 0$, and $g > 0$ on $(0, 1)$. This completes the proof of Lemma 4.1.15. □

*Proof of Lemma 4.1.16.* This follows because $\delta_3(u)$ is concave in $u$, $\delta_3(-1) = c^2(1 - c^{2m-1})p \geqslant 0$, and $\delta_3(-c) = \delta_2(-c) \geqslant 0$, where the latter equality and inequality follow immediately from (4.18) and Lemma 4.1.15, respectively. □

*Proof of Lemma 4.1.17.* It is clear from the second expression for $p_*(m)$ in (4.4) that $p_*(m)$ decreases continuously from $\frac{1}{2}$ to 0 as $m$ increases from 1 to $\infty$. Also, one can verify that $m_*(p_*(m)) = m$ for all $m \geqslant 1$ (here one may use identity (4.19)).

If now $p \in (0, p_*(m))$, then $p = p_*(m_1)$ for some $m_1 > m$, whence $m < m_1 = m_*(p_*(m_1)) = m_*(p)$.

It remains to consider the condition $p \geqslant p_*(m)$. If at that $p > \frac{1}{2}$, then $m \geqslant 1 = m_*(p)$, by (2.3). If, however, $p \in [p_*(m), \frac{1}{2}]$, then $p = p_*(m_1)$ for some $m_1 \in [1, m]$, whence $m \geqslant m_1 = m_*(p_*(m_1)) = m_*(p)$. □

*Proof of Lemma 4.1.18.* Suppose, to the contrary, that statement (III) of Theorem 2.6 is true, while $m < m_*(p)$. Then, by Lemma 4.1.17, one has

$$p < p_* := p_*(m);$$

then, in particular, one has $0 < p < \frac{1}{2}$. Introduce

$$u_p := -\frac{2^{1-\frac{1}{2m}}(m-1)}{(2m-1)(1-p)}.$$

In view of the elementary inequality $p_* \leqslant 1/(2m-1)$ for $m \geqslant 1$ and the condition $p < p_*$, one has

$$p < \frac{1}{2m-1},$$

which implies that

$$-2^{-\frac{1}{2m}} < u_p \leqslant 0.$$

Taking into account these bounds on $p$ and $u_p$ and employing notation introduced in the proof of Lemma 4.1.13, one can see that

$$\partial_\theta \Delta_{1,p,m}(\theta, u_p)|_{\theta=\pi/4} = \frac{2^{1-1/(2m)}(2m-1)(p-p_*)(p-p_{**})}{m\,(1-p)}, \quad \text{where}$$

$$p_{**} := p_{**}(m) := \frac{2m+1+\sqrt{4(m-1)(m+2)+1}}{4(2m-1)} > p_*.$$

Because of the assumption $p < p_*$, it follows that

$$\partial_\theta \Delta_{1,p,m}(\theta, u_p)|_{\theta=\pi/4} > 0.$$

On the other hand,

$$\partial_\theta \Delta_{1,p,m}(\theta, u)|_{\theta=\pi/4} = \frac{m\,2^{1/m-1}}{(1-p)p}\,\partial_\theta \Delta_{p,m}(\theta, -u - 2^{1-1/(2m)}\,p)|_{\theta=\pi/4}$$

for all real $u$; this follows from (4.17), in view of the fact that the derivatives of $\cos^{1-1/m}\theta\,\sin^{1-1/m}\theta$ and $\cos^{1/m}\theta + \sin^{1/m}\theta$ in $\theta$ at $\theta = \pi/4$ are zero. Hence,

$$\partial_\theta \Delta_{p,m}(\theta, t_p)|_{\theta=\pi/4} > 0 \quad \text{for} \quad t_p := -u_p - 2^{1-1/(2m)}\,p.$$

Note also that $\Delta_{p,m}(\pi/4, t) = 0$ for all real $t$. Therefore, $\Delta_{p,m}(\theta, t_p) < 0$ for all $\theta$ in a left neighborhood of $\pi/4$. Now (4.16) implies that $\pi/4$ is not a point of maximum in $\theta$ of $\mathsf{E}\,f_t(\mathrm{BC}_1\,\cos^{1/m}\theta + \mathrm{BC}_2\,\sin^{1/m}\theta)$ for $t = t_p$. Hence, in view of the homogeneity argument used in the proof of Lemma 4.1.13, $\pi/4$ is not a point of maximum in $\theta$ of $\mathsf{E}\,f_t(\mathrm{BS}_1\,\cos^{1/m}\theta + \mathrm{BS}_2\,\sin^{1/m}\theta)$ for $t = t_p/\sqrt{pq}$. But, for any $t \in \mathbb{R}$, one has $f_t \in \mathcal{H}_+^3 \subseteq \mathcal{F}_+^3$ (where the set inclusion follows by Lemma 4.1.2). Thus, one obtains a contradiction with the assumed statement (III) of Theorem 2.6. $\square$

# References

[1] BENTKUS, V. (2001). On measure concentration for separately Lipschitz functions in product spaces. *Israel J. Math.* To appear.

[2] BENTKUS, V. (2002) A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand. *Lithuanian Math. J.* **42**, 262–269. MR1947624

[3] BENTKUS, V. (2003) An inequality for tail probabilities of martingales with differences bounded from one side. *J. Theoret. Probab.* **16**, 161–173. MR1956826

[4] BENTKUS, V. (2004) On Hoeffding's inequalities. *Ann. Probab.* **32**, 1650–1673. MR2060313

[5] BENTKUS, V.; GEUZE, G. D. C.; VAN ZUIJLEN, M. C. A. (2005) Maximal inequalities for super-martingales under a symmetry assumption. *Report No. 0507, Department of Mathematics, Radboud University Nijmegen, The Netherlands.*

[6] BENTKUS, V.; GEUZE, G. D. C.; VAN ZUIJLEN, M. C. A. (2005) Trinomial laws dominating conditionally symmetric martingales. *Report No. 0514, Department of Mathematics, Radboud University Nijmegen, The Netherlands.*

[7] BENTKUS, V.; KALOSHA, N.; VAN ZUIJLEN, M. C. A. (2005) On domination of tail probabilities of (super)martingales: Explicit bounds. *Preprint.*

[8] BERGER, E. (1997) Comparing sums of independent bounded random variables and sums of Bernoulli random variables *Statist. Probab. Lett.* **34**, 251–258. MR1458019

[9] BOBKOV, S. G., GÖTZE, F. AND HOUDRÉ, C. (2001) On Gaussian and Bernoulli covariance representations. *Bernoulli* **7**, 439–451. MR1836739

[10] CHOQUET, G. (1955) Theory of capacities. *Ann. Inst. Fourier, Grenoble* **5**, 131–295. MR0080760

[11] DUFOUR, J.-M. AND HALLIN, M. (1993) Improved Eaton bounds for linear combinations of bounded random variables, with statistical applications. *J. Amer. Statist. Assoc.* **88**, 1026–1033. MR1242946

[12] EATON, M. L. (1970). A note on symmetric Bernoulli random variables. *Ann. Math. Statist.* **41**, 1223–1226. MR268930

[13] EATON, M. L. (1974). A probability inequality for linear combinations of bounded random variables. *Ann. Statist.* **2**, 609–614.

[14] EDELMAN, D. (1990). An inequality of optimal order for the tail probabilities of the $T$ statistic under symmetry. *J. Amer. Statist. Assoc.* **85**, 120–122. MR1137357

[15] EFRON, B. (1969). Student's $t$ test under symmetry conditions. *J. Amer. Statist. Assoc.* **64**, 1278–1302. MR0251826

[16] FELLER, W. (1943). Generalization of a probability limit theorem of Cramér. *Trans. Amer. Math. Soc.* **54**, 361–37. MR0009262

[17] HAAGERUP, U. (1982). The best constants in the Khinchine inequality. *Studia Math.* **70**, 231–283. MR0654838

[18] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30. MR144363

[19] JING, B.-Y.; SHAO, Q.-M.; ZHOU, W. (2004). Saddlepoint approximation for Student's *t*-statistic with no moment conditions. *Ann. Statist.* **32**, 2679–2711. MR2153999

[20] HUNT, G. A. (1955) An inequality in probability theory. *Proc. Amer. Math. Soc.* **6**, 506–510. MR0075470

[21] KHINCHIN, A. (1923). Über dyadische Brüche. *Math. Z.* **18**, 109–116.

[22] LOGAN, B. F.; MALLOWS, C. L.; RICE, S. O.; SHEPP, L. A. (1973). Limit distributions of self-normalized sums. *Ann. Probab.* **1**, 788–809. MR0362449

[23] MARSHALL, A. W. AND OLKIN, I. (1979). *Inequalities: theory of majorization and its applications.* Mathematics in Science and Engineering, 143. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London. MR0552278

[24] MUIRHEAD, R. F. (1903). Some methods applicable to identities and inequalities of symmetric algebraic functions of *n* letters. *Proc. Edinburgh Math. Soc.* **21**, 144–157.

[25] PETROV, V. V. (1965). On the probabilities of large deviations for sums of independent random variables. *Theor. Probability Appl.* **10**, 287–298. MR0185645

[26] PHELPS, R. R. (1966). *Lectures on Choquet's theorem.* D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London. MR0193470

[27] PINELIS, I. (1994). Extremal probabilistic problems and Hotelling's $T^2$ test under a symmetry condition. *Ann. Statist.* **22**, 1, 357–368. MR1272088

[28] PINELIS, I. (1998). Optimal tail comparison based on comparison of moments. *High dimensional probability (Oberwolfach, 1996), 297–314, Progr. Probab., **43**, Birkhäuser, Basel.* MR1652335

[29] PINELIS, I. (1999). Fractional sums and integrals of *r*-concave tails and applications to comparison probability inequalities *Advances in stochastic inequalities (Atlanta, GA, 1997), 149–168, Contemp. Math., **234**, Amer. Math. Soc., Providence, RI.* MR1694770

[30] PINELIS, I. (2003). Dimensionality reduction in extremal problems for moments of linear combinations of vectors with random coefficients. *Stochastic inequalities and applications, 169–185*, Progr. Probab., 56, Birkhäuser, Basel, MR2073433

[31] PINELIS, I. (2005). Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above. *To appear in the Proceedings of the conference on high dimensional probability held in New Mexico, June 2005.* `http://arxiv.org/abs/math.PR/0512301`.

[32] PINELIS, I. (2005). On normal domination of (super)martingales. *Preprint,* `http://arxiv.org/abs/math.PR/0512382`.

[33] PINELIS, I. (2006). On inequalities for sums of bounded random variables. *Preprint,* `http://arxiv.org/abs/math.PR/0603030`.

[34] PINELIS, I. (2006). Toward the best constant factor for the Rademacher-Gaussian tail comparison. *Preprint,* `http://arxiv.org/abs/math.PR/0605340`.

[35] SHAO, Q.-M. (1997). Self-normalized large deviations. *Ann. Probab.* **25**, 285–328. MR1428510

[36] SHORACK, G. R. AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York. MR0838963

[37] WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Teor. Verojatnost. i Primenen.* **5**, 331–335. MR0133849